大局的な構造を考慮した歌詞自動生成システムの提案

渡邉研斗† 松林優一郎† 乾健太郎† 後藤真孝‡

東北大学† 産業技術総合研究所‡
{kento.w, y-matsu, inui}@ecei.tohoku.ac.jp, m.goto@aist.go.jp

1 はじめに

歌詞の創作では，楽曲全体で一貫した内容を表現することが重要である．音楽の創作支援技術として，自動作曲・自動作曲支援システム [1, 2, 3, 4] や歌声合成システム [5, 6, 7, 8] は様々な研究開発がなされており，一般の人々が利用し始めている．自動作詞・作詞支援システムについても様々な研究者が取り組んできたが [9, 10, 11, 12]，従来の歌詞自動生成では，局所的な範囲 (例えば，1 行だけの限られた構造) のみを考慮した生成結果をつぎはぎして楽曲全体の歌詞を生成することが多かった．その結果，楽曲全体の歌詞が与える印象としては表現力に限界があり，A メロや B メロ，サビ等 (これらを以下ブロックと呼ぶ) の楽曲内の構造を十分に考慮しない歌詞しか生成できなかった．

楽曲全体で一貫した内容を表現できるように作詞するためには，単語の依存関係だけでなく，その上位のフレーズ (単語の集合)，行 (フレーズの集合)，ブロック (行の集合)，更には 1 番と 2 番のような「番」(ブロックの集合) との依存関係までを総合的に取り扱う必要がある．そこで本研究では，楽曲全体の歌詞を自動生成することを目標として，

1. 単語からフレーズを生成するモデル
2. フレーズから行を生成するモデル
3. 行からブロックを生成するモデル
4. ブロックから番を生成するモデル

で構成されるモデルを考え，下位の小さな構造のモデルを組み合わせることで上位の大きな構造を生成する手法を提案する．このようにモデルを各階層の構造単位に分けることで，ユーザが求める構造単位での作詞支援が可能となる．例えば，ユーザがフレーズやブロックを指定すれば，その範囲での歌詞の修正や補完を実現しやすくなる．本研究では第一段階として，各構造単位で作詞支援できる柔軟性の高いシステムを開発しつつ，最終的に大局的な構造を考慮した曲全体の歌詞自動生成を目指していく．

提案手法では，単語をつないでフレーズを生成するために，モーラ数制限付き単語 N-gram 言語モデル [1] を使用する．また，テーマに関連する歌詞を生成するために，共起頻度を用いた確率モデルを応用する．加えて，行間の流れを考慮しながらブロックレベルの歌詞を生成するために，Barzilay らが提案したコンテンツモデル [13] を応用する．

[1] モーラは音韻論上の時間的長さを持った音の文節単位である．

本稿では作詞に必要な要素を考慮したモデルと考慮しないモデルでの生成確率を比較し，提案手法の基本的な性能を評価する．また，得られた評価結果から，より質の高い歌詞自動生成に向けた考察をする．

2 関連研究

2.1 作詞支援の関連研究

山本ら [9] は，自然文から歌詞へモーラ数を考慮しながら変換する手法を提案した．この自然文を出発点とする手法は，作者が伝えたい文章を歌詞にできるため有用性が高い作詞支援であるが，自然文がない状態から歌詞を生成することはできない．

阿部ら [10, 11] は，モーラ数，韻，アクセントを考慮した N-gram 言語モデルを用いた作詞支援システムを開発した．吉川ら [12] は隠れマルコフモデルを用いて，歌詞の単語語間の遷移を指定のモーラ数に合わせてモデル化している．しかし，両者は歌詞全体を自動生成するのではなく，作詞支援に特化して，フレーズもしくは行生成のみが可能なモデルを用いていた．

いずれの先行研究でも，行より上位の大きい構造の生成はしておらず，構造の関係や流れを考慮していない問題があった．

2.2 文生成・要約の関連研究

歌詞以外の研究分野では文生成や要約など，様々な研究が行われている．Barzilay ら [13] は，あるトピックの文章から別のトピックの文章への変化を隠れマルコフモデルを用いて学習している．Barzilay らが提案したモデルをコンテンツモデル (CM) と呼び，Ritter[14] や東中ら [15] は CM を利用し，ツイッターのデータを用いた対話モデルの構築に取り組んでいる．本研究では歌詞の流れを考慮するために，CM を利用する．

3 作詞するためのノウハウの調査

本研究では歌詞の階層的な生成モデルを検討する上で，人間が作詞をするプロセスを調査して参考にした．下記の要素が，複数の作詞教本 [16, 17, 18, 19, 20] に述べられていた．

3.1 一貫性のある主張

歌詞には作詞者の主張があり,「誰が誰にいつ何をどのようにして」といった具体的な「ストーリー」を主

張することが多い．また,「愛」や「夢」のようにより大きな視点で，作品全体を表す「テーマ」として主張することもある．

3.2 歌詞とメロディ

歌詞とメロディは密接に関連している．どちらが先に創作されるかは場合によるが，歌詞が先の場合には，後でメロディが付けられるように意識する必要がある．

3.3 歌詞の音楽的構造

歌は通常，上位から順番に 「番」,「ブロック」,「行」,「フレーズ」の構造でできている．作詞する際には，こうした楽曲全体の音楽的な構造を考慮する必要がある．

3.4 音楽的構造間の関係

音楽的構造間には特有の関係がある．1 番と 2 番は通常メロディは類似しており，意味・語感なども類似していて 2 番が 1 番の言い換えと言えることが多い．「A・B メロ」は,「サビ」の内容に共感しやすくするために，起承転結を意識した具体的な背景や補完的な内容が書かれやすい．また，行間の意味的なつながりを考慮する場合がある．例えば，

行間関係の例

帰り道 降り始めた雪 ―――――― 情景
あなたの肩に 触れて 溶けてゆく ――― 情景
今日もまた あっという間だね ――― 感傷
あなたとの週末 終わってしまうの ―― 感傷

岡本真夜 「Everlasting」より抜粋

のように，1，2 行目は客観的な情景について述べているのに対し，3，4 行目は人物の感傷を描いている．このようにブロック内の行には，歌の流れを捉えた意味的な役割が割り当てあられることが多い．

3.5 歌詞内容のバランス

「感情を書くかシチュエーションを書くか」,「主観的か客観的か」,「具体的か抽象的か」,「ポジティブかネガティブか」,「普遍的か斬新的か」などのバランスを，歌の流れに沿って作詞することがある．

3.6 言葉のテクニック

韻を踏む・比喩・ダブルミーニング・二重否定・疑問形・オノマトペ・倒置法・繰り返し・言い換えなどの表現技法を使用することで，歌詞の内容をより強調したり，印象的にできる場合がある．

4 歌詞生成用言語モデル

本稿では前節で述べた作詞に必要な要素を考慮し，先行研究で扱われていない特徴として，歌詞全体の内容が特定のテーマを表現でき，行間の流れを考慮したブロックを生成できる歌詞生成用言語モデルを提案する．

前節で述べたように，作詞する際には「番」や「ブロック」といった音楽的構造を総合的に捉える必要がある．そこで，本研究では以下の仮定をおく．

生成される歌詞の構造の仮定

- 「歌詞」は「番」の連続で生成される．
- 「番」は「ブロック」の連続で生成される．
- 「ブロック」は「行」の連続で生成される．
- 「行」は「フレーズ」の連続で生成される．
- 「フレーズ」は「単語」の連続で生成される．
- 生成される歌詞は与えられたリズムに合う．

つまり，フレーズを生成する枠組みが，歌詞全体を生成する枠組みの一部となっている．また，特定のテーマに関連する歌詞と，行間の意味的な流れを考慮した歌詞を生成するために以下の仮定をおく．

生成される歌詞の内容に関する仮定

- 生成される歌詞は特定のテーマ T について述べている．
- ブロックには行の流れに対応する意味的な役割 C_k(隠れクラス) の遷移 $\mathbf{C} = [C_0, ...C_k, ..., C_q]$ が与えられている．

以上の仮定から，各音楽的構造を生成するモデルの全体図を図 1 に示す．この図のように,「ブロック」,「行」,「フレーズ」の生成部が入れ子になり，歌詞の構造を総合的に考慮したモデルとなっている．また，モーラ数，テーマ，隠れクラスをモデルに組みこみ，歌詞の内容・流れを考慮した生成ができるモデルとなっている．

4.1 使用データ

本研究では，ポピュラー音楽約 2 万 6 千曲の日本語歌詞を，下記の前処理をして用いた．

1. 方言や，アニメキャラクター名を含んだ歌詞を除くために，童謡，民謡，アニメ曲を除く．
2. まずは日本語表現に限定し，英語表現を除外するために，アルファベットを含む行を除く．
3. 形態素解析器 MeCab を歌詞に適用する．

以上の結果得られた 24133 曲，554563 行，57653 単語 (36430 名詞，17812 動詞，2364 形容詞) を使用した．

4.2 言語モデルの説明

図 1 に示すように，テーマとなる単語集合 T とフレーズのモーラ数 N_j の数列 $\mathbf{N} = [N_0, ..., N_j, ..., N_l]$ を，歌詞生成モデルの入力とする．提案するモデルは音楽的構造を階層的に取り扱っているため，上位構造を生成する言語モデルから順に説明する．

まず，ブロック生成モデル $P(B|\mathbf{N}, T, \mathbf{C})$ を式 (1) に示す．ブロック B は行 L_k の連続で生成されるという仮定により，行生成モデル $P(L_k|\mathbf{N}_k, T, C_k)$ の積和でブロックの生成をモデル化する．

$$P(B|\mathbf{N}, T, \mathbf{C}) = \prod_{k=0}^{q} P(L_k|\mathbf{N}_k \subseteq \mathbf{N}, T, C_k \in \mathbf{C}) \qquad (1)$$

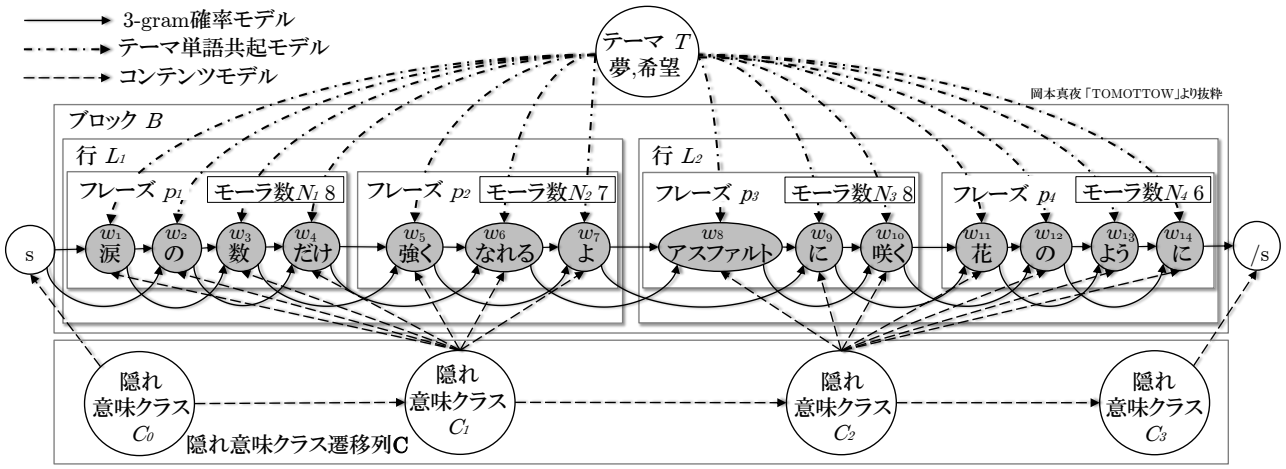

図 1: 歌詞生成モデルの全体像

また，ブロック B には行 L_k の流れに対応した役割 C_k の遷移 $\mathbf{C} = [C_0, ...C_k, ...C_l]$ が与えられていると仮定した．本研究では $\mathbf{C}$ を与えるために，Barzilay ら [13] のコンテンツモデル (CM) を参考にした．Barzilay らは，文章の内容に順序性があるという特徴を利用し，隠れマルコフモデルを用いて文章の構造を推定した．本研究では CM における文章を歌詞に置き換えることで行の流れをモデル化し，状態遷移 $\mathbf{C}$ を推定する．

次に行生成モデル $P(L_k|\mathbf{N}_k, T, C_k)$ を式 (2) に示す．行はフレーズの連続で生成されると仮定したため，フレーズ生成モデル $P(p_j|N_j, T, C_k)$ の積和で行の生成をモデル化する．

$$P(L_k|\mathbf{N}_k, T, C_k) = \prod_{j=0}^{m} P(p_j|N_j \in \mathbf{N}_k, T, C_k) \tag{2}$$

最後に，フレーズ生成モデル $P(p_j|N_j, T, C_k)$ を式 (3) に示す．フレーズは単語の連続で生成されると仮定したため，単語 3-gram 言語モデル $P(w_i|w_{i-1}, w_{i-2}, n_i)$ を用いる．また，テーマ T に関連する単語を生成させるために，テーマ T と歌詞データの単語との共起頻度を使用した単語共起モデル $P(w_i|T, n_i)$ を用いる．加えて，行の流れを考慮した単語を生成させるために，CM によって得られた単語生成確率 $P(w_i|C_k, n_i)$ を用いる．なお，これら単語 3-gram，単語共起モデル，CM の単語生成確率は単語モーラ数 n_i の単語 w_i のみでモデル化する．以上，これら 3 つの言語モデルを線形結合することで，フレーズの生成をモデル化する．

$$\begin{aligned} P(p_j|N_j, T, C_k) = \\ P(\mathbf{n}|N_j) \prod_{i=0}^{l} \big\{ & \alpha P(w_i|w_{i-1}, w_{i-2}, n_i \in \mathbf{n}) \\ & + \beta P(w_i|T, n_i \in \mathbf{n}) + \gamma P(w_i|C_k, n_i \in \mathbf{n}) \big\} \end{aligned} \tag{3}$$

但し $0 \leq \alpha, \beta, \gamma \leq 1$ かつ $\alpha+\beta+\gamma = 1$ であり，α, β, γ の値を変える事で，モデルのバランスを変えることができる．また，入力モーラ数 N_j とはフレーズのモーラ数であり，単語のモーラ数ではない．よって，フレーズのモーラ数 N_j を単語のモーラ数列 $\mathbf{n} = [n_0, ...n_i, ...n_l]$ へマッピングする確率 $P(\mathbf{n}|N_j)$ をフレーズ生成モデルに追加する (式 (4))．

$$P(\mathbf{n}|N_j) = \frac{\text{単語のモーラ数が } \mathbf{n} \text{ であるフレーズの総数}}{\text{モーラ数が } N_j \text{であるフレーズの総数}} \tag{4}$$

4.3 生成方法

生成可能な歌詞の全探索するのは困難であるため，ビームサーチを用いて以下の手順で歌詞を生成する．

1. ビタビアルゴリズム法で状態遷移列 $\mathbf{C}$ を推定する．
2. フレーズ p_j を式 (3) を用いて生成する．
3. フレーズ p_j に続くフレーズ p_{j+1} を生成する．
4. 手順 3 を繰り返して，行 L_k を生成する．
5. 行 L_k に続く行 L_{k+1} を生成する．
6. 手順 5 を繰り返して，ブロック B を生成する．

5 歌詞生成モデルの評価

5.1 評価方法

本研究では，提案したモデルで生成した歌詞が，与えられたテーマ T について述べており，意味的な流れに対応していることを評価したい．しかし，人間の主観による評価の場合，個人によって印象が変わるため，定性的な評価をすることは難しい．そこで本研究では，既存の歌詞には行の流れに対応した意味的な役割 C と，主張したいテーマ T が与えられていると仮定し，既存の歌詞を正解とした定量的な評価を行なう．

まず，テストデータ中の 3 つ以上の行を持つブロックから行 L_{ans} を選び，提案手法で L_{ans} の生成確率を求める．モーラ数列 $\mathbf{N}$ はテストデータの歌詞から自動的に計算したものを入力する．また入力テーマ T となる単語集合は，L_{ans} を除いたブロック内の一般名詞を使用する．意味クラスの遷移列 $\mathbf{C}$ は学習したコンテンツモデルを用い，L_{ans} 以外の歌詞を元にビタビアルゴリズムで推定する．

以上，モーラ数列 $\mathbf{N}$，テーマ T，意味クラスの遷移列 $\mathbf{C}$ を入力し，正解 L_{ans} の生成スコア $Score(L_{ans})$ を計算する (式 (5))．

$$Score(L_{ans}) = \log_{10}\big(P(L_{ans}|\mathbf{N}, T, C \in \mathbf{C})\big) \tag{5}$$

表 1: 各モデルにおける $Score(L_{ans})$ の平均値

使用モデル	α	β	γ	$Score(L_{ans})$
3-gram	1.0	0.0	0.0	-16.62
3-gram+コンテンツモデル (CM)	0.9	0.0	0.1	**-16.40**
	0.8	0.0	0.2	**-16.40**
3-gram+テーマ単語共起モデル (TM)	0.9	0.1	0.0	16.57
	0.8	0.2	0.0	-16.86
3-gram+TM+CM	0.8	0.1	0.1	-16.54

波のリズムは朝を揺らす —— C_3
海はまるで意識の渦 —— C_3
流れは妙に速さを増して —— C_3
非常な程足跡を隠す —— C_6

ACIDMAN 「波，白く」より抜粋

図 2: 歌詞の進行に伴う行の内容の変化を捉えた例

提案したモデルが，テーマと行の流れに沿った歌詞を生成できれば，$Score(L_{ans})$ は大きくなるはずである．

5.2 評価設定

予め，24133 曲の歌詞データを訓練データ:テストデータ=9:1 に分割にし，言語モデルを構築する．

評価ではテーマ単語共起モデル (TM) とコンテンツモデル (CM) を考慮したモデルと考慮していないモデルでの比較を行なうため，式 (3) のパラメータ α, β, γ の値を，表 1 の 6 種類に設定する．

5.3 評価結果と考察

表 1 に各モデルで生成した正解 L_{ans} の生成スコアの平均値を示した．3-gram のみのモデルに比べ，CM を考慮したモデルの $Score$ が高い結果となった．CM の $Score$ が高い歌詞の例を図 2 に示す．図 2 の各行の末尾には CM によって推定された隠れクラス C_k を示した．最初の 3 行は「波」,「海」,「流れ」などの海に関連する単語を使っていることから C_3 は海にまつわる単語のクラスであると推測できる．比べて，最後の 1 行は海とは関連のない単語を使用しているため，別のクラスとなっている．このように CM を用いると，歌詞の進行に合った単語を生成しやすいと考えられる．しかし,「波」や「海」のような特定の内容について明示的に述べている歌詞の他に，言葉の背後に意図が存在する歌詞が多数存在する．本研究で用いた CM は，言葉の背後にある作詞者の意図を考慮していないため，大幅に $Score$ が向上しなかったと考えられる．

また，TM を使用したモデルは，CM や 3-gram を使用したモデルよりも $Score$ 小さくなる結果となった．これは，TM は主張したいテーマを単語で入力し，具体的な主張を考慮していないため，$Score$ が小さくなったと考えられる．

以上，本研究では 3-gram のみを使用したモデルより CM を用いることで $Score$ の向上が得られた．しかし，テーマや行の流れを考慮できる α, β, γ のパラメータのみを使えば良いとは限らず，そのパラメータを楽曲の何処で使用すべきかを推定する必要がある．

6 おわりに

本研究では，単語単位からブロック単位までの音楽的構造を組み合わせ，大局的な構造を考慮した生成モデルを提案した．また，テーマと歌詞内単語との共起頻度や，コンテンツモデルを応用することで行の流れを考慮したモデルを設計できた．

しかし本稿では,「番」や「1 曲」といった更に大きい構造の生成をモデル化できていない．また,「愛」や「夢」と言った単語を入力テーマとしているため，より具体的なシチュエーションを捉えることはできなかった．加えて，提案手法では歌詞の背後にある作者の意図を捉えることができないモデルとなっている．

本研究は 2，3 節で述べたように，まだまだ未発達の領域であり，モデル化すべき歌詞内の言語現象が多く存在する．今後は，提案したモデルの課題点を改善するとともに,「番」や「曲全体」などのより大きな構造や，歌詞内の様々な言語現象のモデル化を目指す．

謝辞

本研究で使用したツールを提供して頂いた東北大学の高瀬翔氏に感謝する．本研究の一部は JST CREST「OngaCREST プロジェクト」の支援を受けた．

参考文献

[1] Katayose H, Hashida M, Poli G D, and Hirata K. On evaluating systems for generating expressive music performance: the rencon experience. In *Journal of New Music Research, Vol.41, No.4*, pp. 299–310, 2012.
[2] 嵯峨山茂樹, 酒向慎司, 堀玄, 深山覚. 確率的手法による歌唱曲の自動作曲. システム制御情報学会誌, Vol.56, No.5, pp. 219–225, 2012.
[3] Fukayama S, Saito D, and Sagayama S. Assistance for novice users on creating songs for japanese lyrics. In *Proc. of ICMC 2012*, 2012.
[4] 自動作曲技術 vocaloducer, 2013. http://jp.yamaha.com/news_release/2013/13102104.html.
[5] 剣持秀紀, 大下隼人. 歌声合成システム vocaloid. Technical report, 情報処理学会 研究報告, 2007.
[6] 剣持秀紀. 歌声合成技術と vocaloid. ヒューマンインタフェース学会誌, Vol.10, Vo.2, pp. 95–98, 2008.
[7] 剣持秀紀. 解説 "歌声合成技術の動向-「初音ミク」を支える技術- ". 日本音響学会誌, Vol.67, No.1, pp. 46–50, 2011.
[8] 後藤真孝. 解説 解説 "「初音ミク」はなぜ注目されているのか ". 電気学会誌, Vol.132, No.9, pp. 630–633, 2012.
[9] 山本貴史, 松原正樹, 斎藤博昭. モーラ数と音節数を考慮した自然文から歌詞への変換. 言語処理学会 第 15 回年次大会 発表論文集, pp. 168–171, 2009.
[10] 阿部ちひろ, 伊藤彰則. 統計的言語モデルを用いた作詞補助システム. Technical report, 2011.
[11] 阿部ちひろ, 伊藤彰則. patissier -アマチュア作詞家のための作詞補助システム-. Technical report, 2012.
[12] 吉川美奈子, 原陽一: 自動作詞システムの開発, 2004. http://www.ipa.go.jp/jinzai/esp/2004mito1/mdata/3-60.html.
[13] Regina Barzilay and Lillian Lee. Catching the drift: Probabilistic content models, with applications to generation and summarization. In *Proceedings of HLT-NAACL*, pp. 113–120, 2004.
[14] Alan Ritter, Colin Cherry, and Bill Dolan. Unsupervised modeling of twitter conversations. In *Proceedings of HLT-NAACL*, pp. 172–180, 2010.
[15] 東中竜一郎, 川前徳章, 貞光九月, 南泰浩, 目黒豊美, 堂坂浩二, 稲垣博人. 2 ツイートを用いた対話モデルの構築. 言語処理学会 第 18 回年次大会 発表論文集, pp. 567–570, 2012.
[16] 遠藤幸三. 作詞本. シンコー Music Entertainment, 2005.
[17] 高田元紀. 作詞のコツがわかる. 中央アート出版社, 2007.
[18] 阿久悠. 作詞入門. 岩波書店, 2009.
[19] 上田起士. よくわかる作詞の教科書. Ymaha Music Media, 2010.
[20] 田口俊. 思いどおりに作詞ができる本. リットーミュージック, 2012.