

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

阿部 ちひろ^{1,a)} 伊藤 彰則¹

概要: 本稿では, Ngram 言語モデルをもとに歌詞候補文を生成する作詞補助システム「pâtissier」への実装を想定した, 歌詞テキストの特徴分析結果を報告する. 作詞においては, 音韻やアクセントなど技巧的な側面の考慮とともに, 楽曲のテーマや歌詞のストーリー設定も重要な要素である. より歌詞らしい候補の生成を目的として, コンテンツ投稿サイト「ピアプロ」に投稿された歌詞テキストを用い, 一般に歌詞らしさと呼ばれる特徴の定量的検討を行った. また, CSJ (日本語話し言葉コーパス) や blog 記事との比較から, 主に使用される単語の違いにより, 歌詞とその他の文章は統計的に区別可能であることが示唆された. さらに, 3 種類のモデルを用いた歌詞生成実験により, それぞれ異なった傾向を持つ文が生成されることが確かめられた.

1. はじめに

近年, DTM(Desktop Music) の普及や動画投稿サイトなどでの楽曲発表の広がりを背景に, 特別な知識を持たない人でも手軽に, 楽曲を制作, 発表することができるようになった. 楽曲制作には作詞・作曲・演奏などの過程があるが, 作詞について学んできた人は少なく, 苦労している人が多い部分である. 作詞においては, 音節数や韻, アクセントの考慮など通常の文章を書くのとは異なる技法が用いられると同時に, 楽曲のイメージやストーリー性, 言葉の美しさを重視する作品も数多く見受けられる. 扱われるテーマや言葉は, 技巧的手法と併せて「歌詞らしさ」を決める重要な要素であると考えられる. しかしながら, 既存の自動作詞システム [1][2] では歌詞全体を自動的に作成するため, ユーザが主体的に好みの言葉を選び, 歌詞に意図を反映させることができない. そこで我々は, 計算機による歌詞生成に技巧的手法と併せて「歌詞らしさ」を反映する作詞補助システム「pâtissier」 [3] を開発している. システムは Ngram 言語モデルをもとに, ユーザの指定した条件を満たす歌詞候補文を生成し, 提示する.

歌詞を扱った研究はさかんに行われており, 社会学的視点から歌詞に表れる時代背景に注目したもの [6] や, 楽曲の意味的解析 [4], 楽曲推薦システム [5] への応用など, 歌詞の持つ特徴の有用性が報告されている. 本稿では, ユー

ザの持つ印象を活かした歌詞生成の実現を目的として, 以下 3 点を中心に調査を行った結果を報告する.

- (1) 歌詞で扱われるテーマについて
 - (2) 歌詞とその他の文を区別する特徴について
 - (3) 歌詞データを用いたモデルによる候補の生成について
- 歌詞データは, コンテンツ投稿サイト「ピアプロ」 [7][8] から収集する. 各歌詞データに付けられたタグは歌詞のテーマを表すものと仮定する. (1) については, 使用頻度の高いタグの調査とタグ別に集めたテキスト群の特徴を解析することで歌詞に表れやすいテーマとそのテーマ特有の特徴量について検討する. (2) については, CSJ (日本語話し言葉コーパス) [9] と blog 記事 [10] から学習した言語モデルと歌詞モデルの違いを確認し, 前者と歌詞を区別している特徴について考察する. また (3) については, それぞれの言語モデルを用いて歌詞の生成を行い, 歌詞らしい文の生成に必要な要素の検討を行う.

2. 作詞補助システム

2.1 システム概要

作詞の手法は, 先に書いた詞に曲を付ける「詞先」と呼ばれる方法と, 出来上がった曲に歌詞を付ける「曲先」と呼ばれる方法に大別される. J-POP に代表されるポピュラーソングでは後者の手法が多く採用されており, 本稿でも曲先での作詞を想定する. 曲先では曲のメロディに歌詞を上手く乗せられるよう単語のモーラ数と韻, さらに音程の上下に合わせたアクセントに気を配る必要がある. そこで, 特別な知識を持たない人が作詞をする際に必要な補助

¹ 東北大学大学院工学研究科
Graduate School of Engineering, Tohoku University
^{a)} abechi@spcom.ecei.tohoku.ac.jp



図 1 作詞補助システム pâtissier

として、

- 曲に合わせられるような音節数の歌詞を提示する
- ユーザーが意図する韻を持つ歌詞を提示する
- より自然にメロディに乗せられるアクセントを持つ歌詞を提示する

という機能が考えられる。このような機能を備えた作詞補助システムの実現のために、本稿では大量の既存文章から統計的に言葉のつながりを学習した Ngram 言語モデルを用いる。Ngram 言語モデルで基本的な単語連鎖を予測し、その上で音節数や韻、アクセントがユーザーの指定条件にある単語列を歌詞候補文として提示する。

モーラ数と母音 音節数と韻の条件を設定するためにモーラ数と母音という要素を導入する。モーラとは一般的に拍と呼ばれるもので、日本語の歌では 1 つの音符に 1 モーラを当てはめる。ユーザーは生成される歌詞候補文のモーラ数を指定することでメロディに合う文を探すことができる。また、モーラがそれぞれの母音を持つかを指定することで生成される文の韻を決めることができる。

システムの使い方 ユーザーはまず、歌詞の先頭にくる単語（入力文と呼ぶ）を入力し、入力文に続く歌詞のモーラ数とそのモーラを持つ母音、アクセントの位置を指定する。システムは入力文に続く単語列を生成し、Ngram 確率と各種条件の充足を考慮してスコアを付け、上位のスコアを持つ単語列を歌詞の候補をして提示する。システムの GUI を図 1 に示す。

2.2 Ngram 言語モデル

歌詞候補文の生成には、Ngram 言語モデルを用いる。Ngram 言語モデルとは、ある単語の生成確率が直前の (N-1) 単語にのみ依存すると近似した言語モデルである。単語の出現確率は学習テキストから統計的手法を用いて推定しているため、各単語の生成確率は学習テキストに依存する。

本稿では文の生成において、式 (1) に示す線形結合により単語 Ngram と品詞 Ngram を併用する。

$$P'(w_i|w_{i-2}w_{i-1}) = \lambda P(w_i|w_{i-2}w_{i-1}) + (1 - \lambda)P(w_i|c_i)P(c_i|c_{i-2}c_{i-1}) \quad (1)$$

ただし、 c_k は単語 w_k の品詞を表すクラス、 $0 \leq \lambda \leq 1$ である。事前実験の結果より、 $\lambda = 0.8$ 、 $P(w_i|c_i) = 1$ とした。

2.3 文の生成

一般に、単語列 $w_1w_2\dots w_n$ の出現確率は 3gram を用いて

$$\log P(w_1w_2\dots w_n) = \log P(w_1) + \log P(w_2|w_1) + \sum_{k=3}^n \log P(w_k|w_{k-2}w_{k-1}) \quad (2)$$

と表される。 w_1w_2 を入力文として、単語列の合計モーラ数がユーザーが指定したモーラ数と等しくなるように単語を選び、文を生成する。文生成アルゴリズムは以下の通りである。

- (1) 単語 w が入力文 h に続く 3gram 確率を求める
- (2) 単語 w を母音条件、アクセント条件とマッチングし、条件不適の場合はペナルティを与える
- (3) 3gram 確率とペナルティから単語列のスコアを決める
- (4) 単語列の末尾 2 単語を新たに h に定め、合計モーラ数が指定モーラ数と等しくなるまで 1~4 を繰り返す

3. 歌詞データの収集

歌詞の特徴を調査するにあたり、まず歌詞データの収集を行った。まとまった量の歌詞テキストを分析することにより、一般的に「歌詞らしさ」と呼ばれる歌詞に多く見られる特徴を統計的に調査し、歌詞らしい文体や単語の使い方、楽曲イメージを考慮したシステム構築へのヒントを探ることを目的とする。

データの収集には、CGM (Consumer Generated Media) 型コンテンツ投稿サイトであるピアプロ [7][8] を用いた。ピアプロには、本システムが対象とするユーザー層である楽曲制作を趣味として行う素人の作品が多く投稿されており、これらの作品はユーザーが付加した属性タグにより検索可能である。

検索タグの準備と歌詞の取得 歌詞データ検索用タグには、サイト全体で使用頻度 15 以上のものから、運営管理用タグ（作曲者募集や処女作）などデータ解析に不向きと考えられるものを人手を除外した 361 タグを採用した。これらのタグは歌詞のテーマを表すものと仮定する。検索タグの例を表 1 に示す。これらの検索タグにより、のべ 9,827 曲の歌詞データ（総形態素数 208 万、語彙数 4 万語）を取得した。ただし、データはタグ別に管理しており、タグ間の重複は確認していない。以降、特に多く（36 曲以上）の歌詞データを集められた 46 タグ、1,669 歌詞を中心にデータ解析を行うこととする。

表 1 検索タグの例

タグ	取得歌詞数
卒業ソング	38
女性視点	29
夏祭り	22
キラキラ	19
壮大	15

表 2 使用頻度の高いタグの分類とその例

テーマ	タグ例
季節	卒業ソング, 桜, クリスマス, 冬
楽曲ジャンル	ギター, ジャズ, 替え歌, 子守唄
感情・状態	切ない, 幸せ, ポジティブ, 笑顔
その他	世界, 人魚, 夜, 夢, 空, 友達

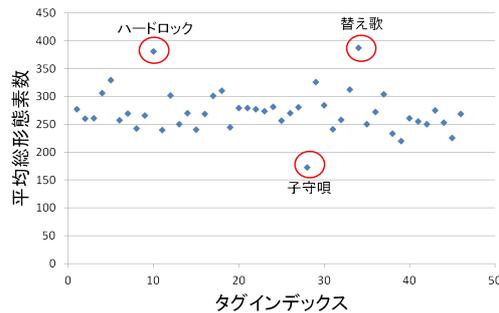


図 2 タグ別 平均総形態素数

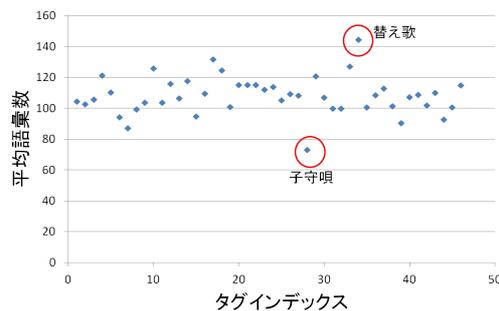


図 3 タグ別 平均語彙数

4. データの解析

4.1 タグに関する特徴

上位 46 タグの品詞内訳は、名詞 44, 形容詞 2 (かわいい, 切ない) であった。使用頻度の高いタグは歌詞で扱われやすいテーマと仮定すると、歌詞のテーマは表 2 のように大別できる。

4.1.1 タグ別 総形態素数と語彙数

各タグ 1 曲当たりの平均総形態素数と平均語彙数を調べた。タグ毎の分布をそれぞれ図 2, 図 3 に示す。

総形態素数はおおよそ一定の範囲内に分布した。外れ値を取ったものを詳しく見てみると、アルファベットを 1 文字 1 形態素として解析したことにより、英単語が多く出現する「ハードロック」では形態素数が他に比べて多くなったことがわかった。同じく形態素数が多い例として「替え歌」では、同じ単語を何度も重ねて連呼したり、長台詞のよう

表 3 特徴的な人称代名詞出現例 (タグ: 演歌)

代名詞	頻度
あなた	79
私	33
貴方	19
あんた	19

表 4 特徴の表れない人称代名詞出現例 (タグ: 切ない)

代名詞	頻度
君	90
僕	71
私	65
あなた	40

表 5 特徴的な名詞出現例 (タグ: クリスマス)

名詞	頻度
クリスマス	75
雪	25
プレゼント	25
今日	24
メリー	22
聖夜	14

表 6 特徴の表れない名詞出現例 (タグ: かわいい)

名詞	頻度
すすめ	20
空	17
今日	16
笑顔	15
夢	15
言葉	12

なものがあったりと、多様な表現が用いられた極端に長い歌詞の曲が数曲見受けられた。反対に形態素数の少ない例「子守唄」では、全体的に簡潔で短い歌詞の曲が多かった。

語彙数についても大きく外れるタグは少なかった。「ハードロック」は日本語の語彙数では他のタグとほとんど変わらず、「替え歌」と「子守唄」はそれぞれ形態素数と同じ理由で他のタグから外れた値を取った。

以上の結果から、歌詞の長さや含まれる単語の数はタグによらず、おおよそ一定の範囲内に集まっているということが分かった。

4.1.2 タグ別 出現単語の観察

まず、人称代名詞に注目してタグ別の単語出現頻度を調べた。特徴的な例と特徴があまり表れなかった例をそれぞれ、表 3, 表 4 に示す。「僕」、「君」、「私」はいずれの歌詞においても頻出の単語であり、特徴の少ないタグではこれらの単語が上位を占めていた。人称代名詞に特徴が現れるのは、かなり限られたタグのみであり、特に曲のジャンルを示すタグで多く見受けられた。

次に、名詞の出現頻度を調べた。特徴的な例と特徴があまり表れなかった例をそれぞれ、表 5, 表 6 に示す。いず

表 7 実験条件 (モデル間)

言語モデル	単語 3gram (バックオフあり)
ディスカウント	Witten-Bell 法
学習データ	piapro : 上位 46 タグ 1,505 歌詞 39 万語 CSJ : 模擬講演から 200 講演 38 万語 blog : 全データからランダムに抽出 9.7 万語
テストデータ	piapro : 残り 113 歌詞 2.9 万語 CSJ : 模擬講演から 19 講演 3.7 万語 blog : 学習データ残り 0.8 万語
評価方法	テストセットパープレキシティ

れのタグにも共通して表れる名詞 (つまり, 日本語そのもの, または歌詞全般の特徴を表す単語) として「こと」, 「もの」, 「世界」, 「今」が多く見受けられた。また, どのタグにおいてもタグと同じ単語は頻出していた。場面や季節を表すもの (クリスマス, 卒業ソング, 戦争) で特徴が出やすかった一方, ジャンル (バラード, ピアノ) を表すタグや感情を表すタグ (かわいい, 切ない) には出現する単語に特徴が出にくい傾向が見受けられた。付加されたタグは一般的に楽曲テーマを表すと考えられるが, 出現する単語とテーマが関連があると一概には言えないようである。

4.2 歌詞とその他の文を区別する特徴

4.2.1 テストセットパープレキシティによる評価

収集した歌詞データセット (piapro と呼ぶ) を CSJ (日本語話し言葉コーパス) [9], 砕けた書き言葉の blog 記事 [10] のデータセットと比較することにより, 両者を区別している特徴を調査する。まず, テストセットパープレキシティを求め, 3 種類のテキストが区別可能であるのかを調べた。パープレキシティとは, 式 (3) に示すように, 単語 1 個が出現する確率の相乗平均の逆数で定義される。このとき, $w_1w_2...w_n$ として, モデル学習に使ったテキストとは別のテキストを用いて算出したものをテストセットパープレキシティ (以下 PP) と呼ぶ。

$$PP = (P(w_1w_2...w_n))^{-\frac{1}{n}} \quad (3)$$

パープレキシティが低い程, テストセット中の単語連鎖の出現確率が高くなる (連鎖する可能性のある単語を絞り込める) ため, テストセットの単語連鎖を上手くモデル化できているということになる。つまり, 学習データとテストデータの種類を入れ替えて評価することで, 各テキストがどのくらい似ているのかを知ることができると考えられる。

表 7 に実験条件を示す。なお, 各種言語モデルは palmkit[12] を用いて作成した。各学習テキストは Sen[13] を用いて形態素解析を行い, 形態素解析器の辞書には UniDic[14][15] を用いた。

図 4 に結果を示す。横軸はテストデータの種類, 凡例はモデルデータの種類をそれぞれ表している。図 4 から 3 種類のテキストはそれぞれ統計的に区別可能であり, piapro が話し言葉とも砕けた書き言葉とも違う特徴を持っている

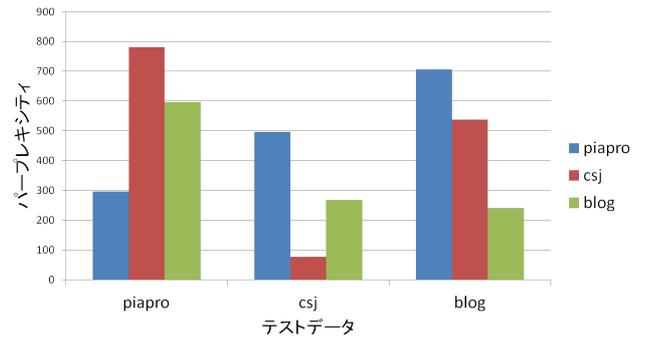


図 4 モデル間 PP

表 8 実験条件 (J-POP 歌詞)

言語モデル	単語 3gram (バックオフあり)
ディスカウント	Witten-Bell 法
語彙数	piapro : 4 万語 CSJ : 4.2 万語
学習データ	piapro : 361 タグ のべ 9,827 歌詞 208 万語 CSJ : 模擬講演 全 1,715 講演 335 万語
テストデータ	嵐 : 4/16 付 検索上位 150 曲 ランキング : 4/16 付 検索上位 100 曲
評価方法	テストセットパープレキシティ

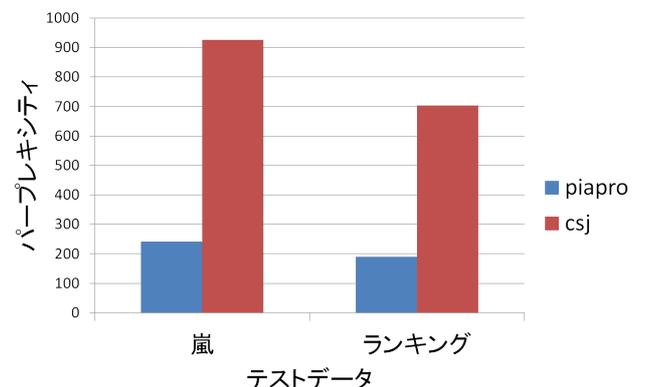


図 5 J-POP 歌詞 PP

ことが示唆された。

次に, piapro が本当に歌詞の特徴を反映しているのかを確かめるため, テストデータとして実際の J-POP の歌詞を用いて同様の実験を行った。歌詞データには歌詞検索サイト「うたまっぷ」[11] の歌手別検索ランキング (歌手名 : 嵐) 上位 150 曲と, 人気歌詞ランキング上位 100 曲を用いた。実験条件を表 8 に示す。

図 5 の結果から, piapro モデルは CSJ モデルに比べて, より実際の歌詞に近いモデルであると考えられることが示された。

4.2.2 未知語の観察

次に, モデル間の違いを明確にするため, テストデータに含まれるが, モデルに含まれない語彙 (未知語) について検

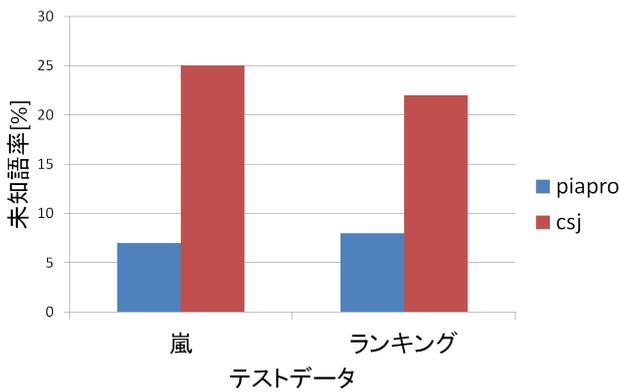


図 6 未知語率

表 9 未知語例

表記のゆれ	キミ+名詞-普通名詞-一般 ココロ+名詞-普通名詞-サ変可能
固有名詞	御茶の水+名詞-固有名詞 マーシャル+名詞-固有名詞
オノマトペ	ニヤニヤ+副詞 ゲラゲラ+副詞
詩的な表現	言の葉+名詞-普通名詞-一般 さんざめく+動詞-一般
若者言葉	マジ+形状詞-一般 バカップル+名詞-普通名詞-一般

表 10 2gram 未知語例

感動詞+何か	ねえ+感動詞 今+名詞-普通名詞 ほら+感動詞 一番+副詞
動詞+名詞	溢れ出す+動詞 光+名詞-普通名詞 舞い戻る+動詞 花びら+名詞-普通名詞
動詞+動詞	喘ぎ+動詞 嘆き+動詞 傷付き+動詞 傷付け+動詞

討した。テストデータに J-POP の歌詞、モデルに piapro と CSJ を用いた場合の未知語率を図 6 に示す。piapro と CSJ では語彙数がほとんど同じであるにも関わらず、未知語率に差が生じた。つまり、歌詞に出現しやすい単語がより piapro に表れているものと考えられる。歌詞に出現するが、CSJ に出現しない単語の具体例を表 9 に示す。CSJ では話し言葉を書き起こしているため、表記のゆれがないのに対して、歌詞では同じ意味の言葉を多様な表記で表している場合が多く見受けられた。また、固有名詞が効果的に使われていたり、オノマトペが多用されるのも歌詞の大きな特徴である。一般的に詩的な表現と言われる様な独特な言葉遣いをする歌詞と、それとは対照的に崩れた日本語、流行言葉が連ねられる歌詞どちらも存在しており、これらは作者の違いに依存するものと考えられる。

さらに、2gram の未知単語連鎖についても同様の検討をおこなったところ、表 10 のような結果が得られた。

表 11 生成文の比較 (入力文: 恋愛は, モーラ数: 4, 母音: - a i -)

piapro	CSJ	blog
悲しみ	こなして	暇人
哀しみ	いらして	ちなみに
七色	果たして	願いし
世界を	来たりと	都会の
隣を	いらした	柳と

表 12 生成文の比較 (入力文: 恋が, モーラ数: 10)

piapro	CSJ	blog
滲み出してこぼれた	分からないんですけど	したいのかわからない
始まった物語	好きだったんですけど	したい人並み程度
滲みだしてこぼれて	あったんですけどでも	したいのかわかりませ
始まった前世が	あるんですけどっていう	したい人並みっから
始まった五月雨の	始まったんですけど	したいのかわかります

表 13 卒業ソング+桜 (入力文: 恋が, モーラ数: 10)

さくころ巡り巡っ
さくころ翼広げ
さくころね想い出す
さくころねめくるめく
さくころねコトバない

4.3 歌詞モデルによる文の生成

3 種類の言語モデルを用いて文の生成を行い、それぞれの違いを検討した。piapro モデルでは、全歌詞データを用いた言語モデルとは別に、使用頻度上位 46 タグについてタグ別モデルも作成した。タグ別モデルは各タグ 36 または 37 歌詞のテキスト群を学習データとし、1 タグあたりの平均形態素数は 0.9 万語であった。

4.3.1 言語モデル間の違い

歌詞生成条件 (入力文, モーラ数, 母音条件) 150 セットについて、3 種類の言語モデルそれぞれで文の生成を行い、その違いを観察した。生成された文の例を表 11 に示す。短いモーラ数の条件では生成される文が大きく違うと言うには至らない結果となったが、piapro モデルは歌詞らしくない文を除外できていると言える。

次に、モーラ数を 10 モーラに増やして文の生成を行った。結果を表 12 に示す。ある程度の長さの文を生成してみると、各言語モデルの特徴が出やすいことが明らかになった。

4.3.2 タグ別言語モデルの線形結合

次に、タグ別に作成した言語モデルを用いて文を生成し、結果を観察した。類似する 2 種類の言語モデルを手で選び、式 (4) に示すように線形結合を用いて組み合わせて使用した。例として「卒業ソング」と「桜」の結果を表 13 に示す。

$$P(w_i|w_{i-2}w_{i-1}) = 0.5P_1(w_i|w_{i-2}w_{i-1}) + 0.5P_2(w_i|w_{i-2}w_{i-1}) \quad (4)$$

歌詞らしい印象を受ける文が生成されたが、卒業や桜を

イメージさせる歌詞の生成とまでは言えない結果であった。どのようにモデルを組み合わせるのが適切であるのか、さらなる調査が必要である。

5. まとめと今後の課題

本稿では、歌詞投稿サイトピアプロを用いて歌詞データの収集を行い、作詞補助システムへの実装を目的とした歌詞テキストの分析を行った。これにより、以下のような結果を得た。

- 歌詞の長さや出現する単語の種類はタグの種類（楽曲テーマ）によらず、おおよそ一定の範囲内に分布する
- 特徴的な人称代名詞は特定の楽曲ジャンルのタグを持つ歌詞に表れることもあるが、一般的には「僕」「君」「私」が頻出する
- タグを連想するような特徴的な名詞は、場面や季節を表すタグに多く出現するが、ジャンルや感情を表すものでは出現しにくい
- 歌詞とそれ以外の文は統計的に区別可能であり、主に出現する単語に違いがある
- 歌詞データを用いた言語モデルを用いても必ずしも、歌詞らしい文が生成できるわけではないが、歌詞らしくない単語をある程度除外することができる

今後は歌詞モデルを用いて、より歌詞らしい文を生成可能なアルゴリズムの検討を行っていきたい。

謝辞 歌詞データの収集を快諾して下さった、クリプトン・フューチャー・メディア株式会社に感謝する。

参考文献

- [1] 伊藤雅光：ユーミンの言語学 (1)-(46), 日本語学 16(4)-20(8) 連載, 明治書院, (1997-2000).
- [2] HRG Oliveira, Amilcar Cardoso, Francisco Camara Pereira : Tra-la-Lyrics: An approach to generate text based on rhythm, Proc. Int. Joint Workshop on Computational Creativity, (2007).
- [3] 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, 情報処理学会研究報告, Vol.2011-MUS-91 No.9, (2011).
- [4] Beth Logan, Andrew Kositsky, Pedro Moreno : Semantic Analysis of Song Lyrics, International Conference on Multimedia and Expo, (2004).
- [5] 梶 克彦, 平田 圭二, 長尾 確：状況と嗜好に関するアンケートに基づくオンライン楽曲推薦システム, 情報処理学会研究報告. MUS, vol.2004-MUS-58, pp.33-38, (2004).
- [6] 久保 正敏：ニューミュージックに見る恋愛風景, 情報処理学会研究報告. 人文科学とコンピュータ研究会, Vol.1995, No.14, pp.49-57, (1995).
- [7] PIAPRO (ピアプロ), 入手先 (<http://piapro.jp/>).
- [8] 伊藤博之：初音ミク as an interface, 情報処理, vol. 53, no. 5, pp. 477-482, (2012).
- [9] 独立行政法人国立国語研究所, 日本語話し言葉コーパス,(2004).
- [10] にしのそらにみかづき,
入手先 (<http://crescent213.blog12.fc2.com/>).

- [11] うたまっぷ, 入手先 (<http://www.utamap.com/>).
- [12] 伊藤彰則, 好田正紀, 単語およびクラス n-gram 作成のためのツールキット, 信学技報, NLC 100(521), pp.67-72, (2000).
- [13] sen project, 入手先 (<http://sen.dev.java.net/>).
- [14] 伝康晴, 小木曾智信, 小椋秀樹, 山田篤, 峯松信明, 内元清貴, 小磯花絵：コーパス日本語学のための言語資源：形態素解析用電子化辞書の開発とその応用, 日本語科学, 22号 pp.101-122, (2007).
- [15] 伝康晴, 宇津呂 武仁, 山田 篤, 浅原 正幸, 松本 裕治：話し言葉研究に適した電子化辞書の設計, 第2回話し言葉の科学と工学ワークショップ講演予稿集, pp. 39-46, (2002).