

統計的言語モデルを用いた作詞補助システム

阿部 ちひろ^{†1} 伊藤 彰 則^{†1}

本稿では、音節数と韻に着目した作詞支援システムを提案する。システムは Ngram 言語モデルをもとに、ユーザの指定した音節数と韻の条件を満たす歌詞候補文を生成し、提示する。ユーザはシステムを辞書のように使い、提示文から主体的に言葉を選び作詞を進めることができる。我々は GUI を備えた作詞補助システムを実装し、提示文とシステムの主観評価実験を行った。

A Lyrics Writing Support System Using a Statistical Language Model

CHIHIRO ABE^{†1} and AKINORI ITO^{†1}

In this paper, we propose a lyrics writing support system focused on the number of syllables and rhyme. The system generates candidate sentences that satisfy user-specified conditions based on Ngram, and presents them. A user can use the system like a dictionary, and write lyrics by choosing presented words. We have implemented a system with GUI, and subjective evaluations of the statements and proposed system were conducted.

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、個人による楽曲制作が増加している。デジタル技術の発展は、特別な知識を持たない人であっても趣味として楽曲を制作し、自ら発信することを可能にした。そして、誰もが楽曲を聴く喜びだけではなく、アーティストとして楽曲を制作、発表する喜びを享受でき

る。このような傾向は今後さらに顕著になっていくものと考えられる。楽曲制作には作詞・作曲・演奏などの過程があるが、作曲や演奏に比べて作詞について学んできた人は少なく、楽曲制作を行う中で苦勞している人が多い部分である。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、ユーザが自分の好みの言葉を主体的に選ぶことはできない。そこで本研究では、個人による作詞を補助することを目的として単なる自動作詞ではなくユーザが主体的に歌詞を生成、選択出来る作詞補助システムの検討を行う。

2. 作詞補助システム

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、本研究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に気を配る必要があるため、特別な知識を持たない人が作詞をする際に必要な補助として、

- 曲のメロディに合わせられるような歌詞を提示する
- ユーザーが意図する韻を持つ歌詞を提示する

という基本的な2つの機能が考えられる。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、本研究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がユーザの指定条件にあう単語列を歌詞候補文として提示する。歌詞生成に適した言語モデルと文生成アルゴリズムに焦点を当て検討を行い、ユーザの創作意欲を引き出す魅力的な候補文の提示を目指す。

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、モーラがそれぞれど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。例えば、3モーラ末尾「あ」と条件指定した場合は「みんな」、「かぼちゃ」といった単語が候補として提示される。

3. 言語モデルの作成

コンピュータを用いて文を生成するためには、何らかの方法で人間の作る文章の単語連鎖

^{†1} 東北大学大学院工学研究科

Graduate School of Engineering, Tohoku University

をモデル化する必要がある。単語の繋がり方は無限に考えられるため、モデル化には大量の既存文章を学習することにより自動的にモデルを作成する統計的言語モデルの手法がしばしば用いられる。本研究では、統計的言語モデルの中でも広く用いられている Ngram 言語モデルを用いる。

3.1 Ngram 言語モデル

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。学習テキスト中での単語の繋がり方を学習することにより作成されるため、各単語の生成確率は学習テキストに依存する。

3.2 単語 Ngram と品詞 Ngram

単語自体のつながりをモデル化したものを単語 Ngram と呼び、単語の品詞のみを考慮して品詞同士のつながりをモデル化したものを品詞 Ngram と呼ぶ。単語 Ngram では、学習テキスト中に実際に出現した単語の連鎖に高い Ngram 確率が付くため、学習に用いた文に大きく依存する。つまり、歌詞候補文生成の際に学習テキスト中の文と似通った文が生成されやすくなる。品詞 Ngram では学習テキストに出現した単語は品詞が同じならば全て同一のものと思なすため、実際には出現していない単語連鎖にも高い確率を与えることができる。候補文の種類を増やせる一方、学習テキストに実在しない単語連鎖が生まれるため日本語として不自然な候補文が生成される可能性もある。

単語 Ngram (次の 2 文を区別する)

- 今日 は 晴れ
今日 名詞-副詞可能 + は 助詞-係助詞 + 晴れ 名詞-一般
- 今日 は 雨
今日 名詞-副詞可能 + は 助詞-係助詞 + 雨 名詞-一般

品詞 Ngram (次の 2 文は同じものと見なす)

- 今日 は 晴れ
名詞-副詞可能 + 助詞-係助詞 + 名詞-一般
- 今日 は 大学
名詞-副詞可能 + 助詞-係助詞 + 名詞-一般

表 1 学習テキストの詳細
Table 1 detail of corpora

学習テキスト	総形態素数	異なり単語数
ブログ記事	11.6 万	1 万語
CSJ 模擬講演前半	162 万	3 万語
CSJ 模擬講演後半	162 万	3 万語

本研究では、式 (1) に示す線形結合により 2 つの言語モデルを併用した。

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

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。

3.3 言語モデルの作成

本研究では、Palmkit³⁾を用いて単語 3gram と品詞 3gram を作成した。学習テキストには、砕けた書き言葉のブログ記事⁴⁾と国立国語研究所によって開発された「日本語話し言葉コーパス」(Corpus of Spontaneous Japanese:以下 CSJ と略称)⁵⁾の模擬講演を用いた。さらに模擬講演、全 1,715 講演を識別番号の前半、後半で 2 つに分け、blog 記事を合わせ 3 種類の学習テキストとした。表 1 に、各テキストの詳細を示す。また、言語モデルを作成するにあたり学習テキストに対して、以下に示すような句読点と記号の除去の前処理を行った。

処理前

ちょうど見頃で、ちょーきれいだった(人´▽`)。☆。.:*。°

処理後

ちょうど見頃でちょーきれいだった

このような処理を行うことで、歌詞としては有意性に欠ける単語連鎖を除くことができる。前処理を施した後、Sen⁶⁾を用いて学習テキストの形態素解析を行った。Sen の辞書には日本語辞書 ipadic ver2.6.3⁷⁾を用いた。形態素解析によりテキスト中の各単語に読みと品詞の情報を付加し、形態素解析器の辞書にない単語は全て未知語 (UNK)1 単語に置換した。さらに、形態素解析により得られた情報を元に、学習テキスト中に出現した全単語とそれぞれのモーラ数をリスト化した。歌詞候補文はこの単語リスト内の単語を組み合わせることにより生成される。

4. 文生成アルゴリズムの概要と文の生成

システム全体では式 (2) に示すように、ある文 h に続く単語列 W のうちモーラ数 m で確率最大のを歌詞候補文として提示する。

$$\hat{W} = \arg \max_{|W|=m} P(W|h) \quad (2)$$

ユーザは、生成される候補文の直前に来る単語列（入力文と呼ぶ）とモーラ数、母音の条件を指定する。例えば、入力文：今日は、モーラ数：5、母音：a - a - a、のように条件を指定すると、入力文「今日は」に続き、モーラ数と母音の条件を満たす候補文として「暑かった」や「寒かった」といった文が歌詞候補文として提示される。

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

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

と表される。両辺対数を取ると

$$\log P(w_1w_2\dots w_n) = \log P(w_1) + \log P(w_2|w_1) + \sum_{k=3}^n \log P(w_k|w_{k-2}\dots w_{k-1}) \quad (4)$$

となる。ここで、 w_1w_2 を入力文と考えると、 $\log P(w_1) + \log P(w_2|w_1)$ は定数であるから

$$\sum_{k=3}^n \log P(w_k|w_{k-2}\dots w_{k-1}) \quad (5)$$

が大きくなる単語列 $w_3\dots w_n$ を歌詞候補文として提示すればよい。ただし、単語列の合計モーラ数がユーザが指定したモーラ数と等しくなるように単語を選ぶ。

4.1 文生成アルゴリズム

歌詞候補文の生成は次のような手順で行う。

- step1** リスト内の単語 w が入力文 h に続く 3gram 確率を求める
- step2** 単語 w を母音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
- step3** 3gram 確率とペナルティから単語列のスコアを決める
- step4** 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 step1~4 を繰り返す

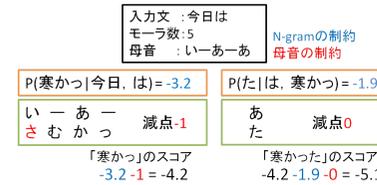


図 1 スコア付けの例
Fig. 1 An example of scoring

表 2 x の値による候補文の違いの例
(入力文「雨の」、母音条件「a-o-」)

Table 2 Difference of sentences by λ value

$\lambda = 1$	$\lambda = 0.8$
話を	内容を
話に	タイトル
話し	会場
話で	誕生
話も	買い物

母音条件を完全に満たす単語のみを文生成に用いることができれば望ましいが、リスト内の限られた単語の組み合わせで自然な文を生成するためには Ngram 確率による基本的な単語連鎖の制約もある程度保つ必要がある。そのため、まず 3gram 確率に基づき単語列を作り、母音条件に適合しない場合に減点ペナルティを与えることで単語を 1 つ選ぶ度に単語列にスコア付けを行う (図 1)。減点を大きくすることで母音条件の制約を強めることができる。

提示文の探索には、ビームサーチを用いる。ビームサーチでは最適解が得られる保証はないが、候補文は必ずしも最適解である必要はない。そのため、計算量と候補文の出来との兼ね合いが取れるビーム幅を用いればよい。リストから単語を 1 つ選び単語列を展開する度にスコア順に単語列をソートし、上位ビーム幅個の単語列のみをさらに探索していく。展開中の単語列の合計モーラ数が指定モーラ数に等しくなった場合はその時点で探索は打ち切り、スコアを保持する。全ての単語列の探索が終わるまで保持したスコアを含めて単語列のソートを行い、最終的に残ったスコア上位の単語列を提示文として採用する。

4.2 文の生成

作成した言語モデルと提案アルゴリズムを用いて実際に文を生成し、言語モデルや各種パラメータの値について検討した。まず、式 (1) λ の値により生成される文にどのような違いが見られるかを検討した。最も顕著な違いは $\lambda = 1$ の場合、母音条件不適合の文が 60 文中 26 文であったのが、 $\lambda = 0.8$ では 0 文であったことである。ここから、品詞 Ngram の比率を調整することで、条件を満たさない文を大幅に減らすことができると考えられる。また、 $\lambda = 0.8$ の場合では表 2 に示すように文に用いられる単語の種類を増やすこともできた。

次に、候補文探索のビーム幅について検討した。図 2、図 3 は、縦軸に対数確率、横軸にモーラ数を取り、ビーム幅ごとに上位 10 候補文の出現確率の平均値をプロットした結果で

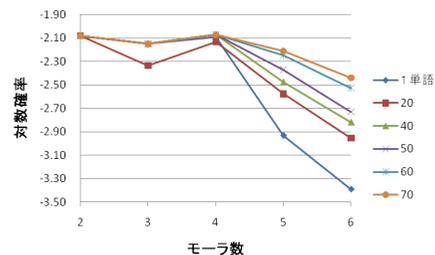


図2 ビーム幅の検討
(入力文「今日は」, 母音条件なし)
Fig.2 Examination of beam width

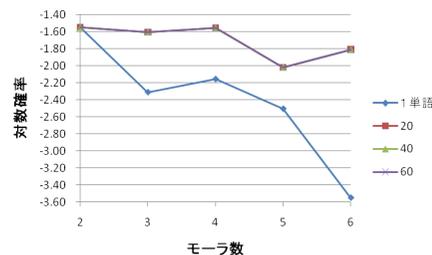


図3 ビーム幅の検討
(入力文「ぼくは」, 母音条件なし)
Fig.3 Examination of beam width

ある。4モーラまではビーム幅の違いによる確率の大きな差は見られなかったが、モーラ数が増えるに従ってビーム幅を大きく取った方がよりよい確率の候補文を生成できることが分かった。しかし、ビーム幅70以上では上位10候補文に変化が見られなくなり、入力文「ぼくは」ではビーム幅20で変化が見られなくなった。適切なビーム幅は生成条件によって大きく異なることが分かった。

最後に、母音条件を満たさない単語に対する減点について検討した。表3に示すように、減点1の場合は母音条件を満たさないものがある一方、減点2の場合は母音条件を満たすために自然な単語連鎖が崩れてしまっていることが見て取れた。

5. 評価実験

5.1 評価実験1

評価実験1ではシステムが提示した歌詞候補文を5人の評価者が主観評価した結果から、生成条件の違いによる評価値の差と評価者間の評価値のばらつきについて検討した。4つの言語モデル条件(表4)それぞれに対して入力文4種(表5)×母音条件3種(図4) = 12条件を指定し、各条件でのスコア上位5文を評価用の文として採用した(計240文)。

評価は、文が自然であるか、文が詩的であるかを基準に、2点:よい、1点:ふつう、0点:文として不自然、の3段階で行い、各文に2点×5人 = 10点満点の主観評価値を付けた。主観評価点の例としては、10点満点「乾いた感情」、7点「雨の誕生」、0点「今日はり過ぎしていた」などが挙げられる。ここで、母音条件の充足を考慮した評価値として

$$\text{母音正解率} = \frac{\text{満たせた条件の数}}{\text{条件の数}} \quad (6)$$

表3 減点の検討

(入力文「ぼくが」, 母音条件「uaua」)

Table 3 Examination of penalty

減点1	減点2
手助けし	スさすがに
いなくなっ	スが上手く
いなくなる	スカ普段
スお名前	スは素直
いたつもり	スは上手く

表4 言語モデル条件

Table 4 Conditions of language model

条件名称	学習テキスト	λの値
blog (単語)	blog 記事	1
blog (単語+品詞)	blog 記事	0.8
csjA	CSJ 模擬講演前半	0.8
csjB	CSJ 模擬講演後半	0.8

条件1: 2モーラ「-u」
条件2: 4モーラ「a-o-」
条件3: 6モーラ「iu-a」

図4 モーラ数と母音の条件

Fig.4 Conditions of mora and vowel

入力文1:「今日は」
入力文2:「先生の」
入力文3:「雨の」
入力文4:「乾いた」

図5 評価実験1の入力文

Fig.5 Input sentences

を定義する。さらに、主観評価値と母音正解率を用いて最終的な文の評価値を

$$\text{評価値} = \text{主観評価値} \times \text{母音正解率} \quad (7)$$

のように定義する。

実験結果と考察

言語モデル条件 blog (単語) と blog (単語+品詞) を比較し、品詞 Ngram の比率により提示文の評価値にどのような違いが見られるかを検討した。図6に示すように2つの条件間に評価値の大きな差は見られなかった。また、評価文を評価値上位、中位、下位、各20文に分けた各層の平均評価値を図7に示す。ここでも両者ともほぼ同じ値が得られた。次に、各層における評価値の標準偏差から評価者間の点数のばらつきを検討した。図8に示すように、中位の標準偏差において両者に差が見られた。つまり、blog で5点前後の評価値を得た文は評価者によって評価が分かれたと言える。例えば「今日はゆず」(評価値5.0, 標準偏差1.0) というような文が挙げられる。これは品詞 Ngram の比率を調整することに

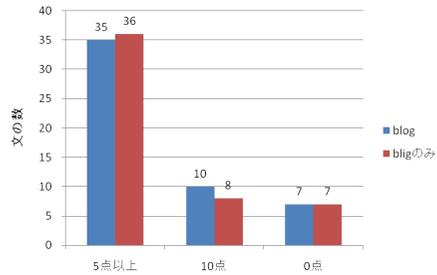


図 6 blog (単語) と blog (単語+品詞) 評価値の分布
Fig. 6 Evaluation value of blog(words) and blog(words+word class)

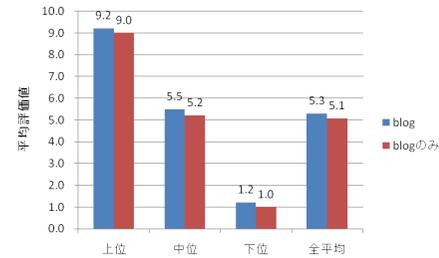


図 7 blog (単語) と blog (単語+品詞) 平均評価値
Fig. 7 Average evaluation value of blog(words) and blog(words+word class)

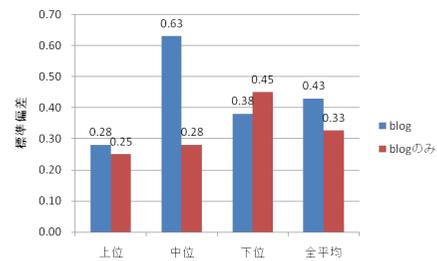


図 8 blog (単語) と blog (単語+品詞) 標準偏差
Fig. 8 Standard deviation of blog(words) and blog(words+word class)

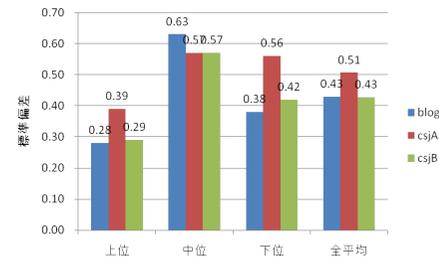


図 9 3条件での標準偏差
Fig. 9 Standard deviation of 3 conditions

より、品詞間の繋がりに基づいて確率が与えられ、母音条件を満たす文が学習テキスト中に出現しない場合でも適切な文を生成可能になった一方、学習テキストに実際には出現しない単語連鎖が生成されたために、評価者間で評価にばらつきが出る奇抜な文が多くなったとも考えられる。

さらに、言語モデル条件 blog (単語+品詞)、csjA、csjB を比較し、学習テキストの規模と内容の違いにより提示文の評価値にどのような違いが見られるかを検討した。csjA と csjB は blog に比べて 5 点以上の評価値を獲得した文が多かった。また、10 点満点と 0 点を取った文はいずれも csjA では若干少なく、blog と csjB ではほぼ同数であった。前者の原因としては、blog は学習テキストの総形態素数が少ないのに加えて、砕けた文体であるため形態素解析が上手く処理されず不自然な単語連鎖がモデル化されてしまい、全体的に低い評価値が付いたと推測される。また後者については、csjA と csjB では規模と文体はほぼ同じであるにもかかわらず評価値に差が出たことから、極端に評価が良い文と悪い文の生成には学習テキストの規模や文体よりも、その内容が大きく関わっていることが推測される。上、中、下位各層の平均評価値は、上位 20 文においてはいずれの条件でも評価値に大きな差はなく平均 9 点前後であったが、中下位においては条件により平均評価値にばらつきがあった。つまり、どの言語モデルを用いてもある程度の数は高評価値の文を生成する事が可能であるが、学習テキストの規模や内容によって中下位の評価値が上下することが分かった。図 9 に示すように、評価値の標準偏差は csjA で大きな値を取っている。評価が分かれる様な変わった文の生成頻度は言語モデルの規模ではなく、学習に用いた学習テキストの内容に強く依存することが考えられる。

以上から、言語モデルの違いが生成文の評価に与える影響は、規模の違いによるものと学習テキストの内容の違いによるものの 2 つに大別されることが分かった。言語モデルの規模が大きいほど下位の評価値が底上げされ、全体として評価値が良くなる傾向が見られた。一方、評価者により評価がばらつく文や、極端に評価が偏る文の生成については学習テキストの規模よりも出現した単語や言い回しの影響をより強く受けることが分かった。

5.2 評価実験 2

評価実験 2 では提案アルゴリズムを実装した作詞補助システムを構築し、システムの使用感の評価を行った。5 名の評価者に図 10 のようなパズル式の作文課題をシステムあり・なし、それぞれの条件で取り組んでもらいシステムの有無が課題に与える影響や、システム自体の使い心地について自由記述式のアンケートを行った。マスの数は音節数、マスの母音指定箇所は韻の制約をそれぞれ意図して出題している。実験に用いたシステムの外観とシス

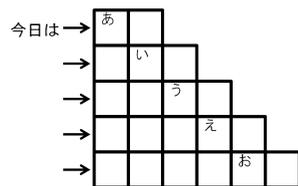


図 10 回答用紙
Fig.10 Answer sheet



図 11 ユーザーインターフェイス
Fig.11 User interface

表 5 評価実験 2 における各種条件設定

Table 5 Conditions for evaluation experiment2

言語モデル	λ の値	ビーム幅	ペナルティ
blog (単語+品詞)	0.8	70	-1

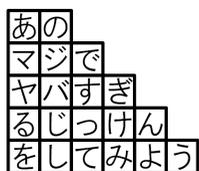


図 12 回答例 システムなし
Fig.12 Example answer without System



図 13 回答例 システムあり
Fig.13 Example answer with System

テムの各種条件をそれぞれ表 5, 図 11 に示す.

実験結果と考察

ある被験者によるシステムあり・なしにおける回答例をそれぞれ図 12, 図 13 に示す. 課題にかかった時間はシステムありの方が長かったが, いずれの場合でも母音条件を満たしており制限時間内に完成できた. 評価者によるアンケート結果を以下に示す.

- 楽しかった
- 想像以上に使えると思った
- 候補文を文末で終わるように指定出来ると思った
- 思ったよりも無難な候補が多かった, 個性が足りない

アンケート結果からユーザはシステムにインパクトのある個性的な単語, 自分では思いつか

ない様な文章を求めていることが分かった.

6. おわりに

本研究では, 個人による作詞を補助することを目的とした作詞補助システムの構築のために, Ngram 言語モデルと文生成アルゴリズムの検討を行った. 品詞 Ngram の比率を調整することにより, システム提示文の条件充足率とバリエーションが向上することが示された. また実際にシステムを構築し, 提示文と GUI の使い心地について評価実験を行った. 評価者に楽しんで使ってもらい, 作詞のアイデアを膨らませるという本来の目的をおおむね達成することができた.

今後は碎けた表現における形態素解析精度の向上と, ユーザ自由度の高い直感的で使いやすい GUI の検討が課題である.

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 他: Tra-la-Lyrics: An approach to generate text based on rhythm, Proc. Int. Joint Workshop on Computational Creativity, (2007).
- 3) 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, 電子情報通信学会技術研究報告, NLC 100(521), pp.67-72, (2000).
- 4) にしのそらにみかづき, 入手先(<http://crescent213.blog12.fc2.com/>).
- 5) 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).
- 6) sen project, 入手先(<http://sen.dev.java.net/>).
- 7) 浅原正幸, 松本裕治, ipadic version 2.6.3