

音と意味の関連性計量による韻表現形成システム

木村 龍輝[†] 清木 康[‡]

[†]慶應義塾大学総合政策学部 〒252-0882 神奈川県藤沢市遠藤 5322

E-mail: [†] {s16278rk, kiyoki}@sfc.keio.ac.jp

あらまし 作詞において歌詞の意味は重要な要素であるが、言葉の響きも作品のクオリティを左右する重要な要素である。しかしながら言葉の響きを大きく左右する「韻を踏む」という行為は英語等に比べて日本語では難しく、経験的習熟を要する。本研究では (1)韻表現特有の音の崩し方 (2)異なる品詞の組み合わせ (3)意味的関連性 の3点に着目して、音と意味を関連性計量により適切な韻表現を形成するシステムを示す。

キーワード テキストデータ処理・テキストデータベース、情報推薦・パーソナライゼーション、作詞支援

1. はじめに

「韻を踏む(押韻)」という行為は、歌詞の中で同じような響きのフレーズを効果的に配置することによって、歌声のリズムや響きを整えることを指す。そのため歌において「韻を踏む」という行為は目立ちこそはしないものの楽曲の最終的な品質に非常に大きく関わるものとなる。

しかし、日本語で韻を踏むには経験的習熟を必要とするため、初心者には上質な韻を踏むことは難しい。英語では名詞が「tion」「sion」という形で終わることも多いため韻が踏みやすいが、日本語にはそのような形式があることは少なく、英語等より押韻の難易度が高い。

押韻を補助する韻表現の提案システムはすでに存在するが、それらは基本的には母音の一致率についてのみ考慮したものとなっている。ただ母音を一致させるだけであると非常に堅苦しい韻となり、提案可能な単語も限られてしまう。

本稿にて提案するシステムは、自然な韻表現をする際の「崩し」や、助動詞の考慮、頭韻/脚韻の考慮など、実用的な韻表現を生成するための技術の仕組みの組み合わせを解説していく。意味も考慮していくことによってより適切な韻表現をすばやく得ることを可能にする。

2. 構成

本稿のシステムは「単語データベース」と「検索機能」の2つから構成される。

本システムは、「単語データベース」を構成し、単語と読みが結び付けて格納している。「検索機能」では韻として成立する単語をデータベースから抽出する機能を指す。これらを組み合わせることによって、様々な単語データベースを音という観点で結びつけることが可能となり、韻を踏む機能を実現する。

3. 単語データベース

単語データベースには文書ファイルから形態素解析によって抽出した単語を「読み方」、「品詞」とともに格納する。

本システムで扱う品詞は「名詞」と「動詞」、「形容詞」のみとする。一般的に踏まれる韻はこの3つで、これらは単一でも意味が通るからである。

4. 母音の音のゆらぎ

本稿のシステムは実用的な韻を提案するために、母音の近さを考慮する。

例えば「望遠鏡(ouenou)」という単語の場合、「桃源郷(ouenou)」という単語と母音が完全に一致しているが、このような母音の完全一致は文字数が増えれば増えるほど限られていき、意味を考慮するとあまり現実的ではない単語しか残らなくなってしまふ。しかし実際は母音を完全一致させる必要はなく、発音の仕方の工夫で多少母音に違いがあっても問題がない。今回の例でいうと、「ouenoo」「oouenou」などがそれに当たる。

本システムにおいてはこれらの母音のゆらぎを考慮して選択肢を増やすことによって実用的で意味の通る韻表現を提案していく。

本稿では対応表として表 1 に示すものを用いる。横 1 行のものはすべて同様に置換できるものとして扱っていく。

表 1 母音のゆらぎの対応表

っ	ん	
う	ん	
おー	おう	おお
あー	あう	
えー	えい	ええ
あい	あ	

5. 助動詞の考慮

実際に韻が踏まれる際には単語単位ではなく、助動詞とセットで韻が踏まれることが多い。例えば「サバナ(aana)」という単語に対して「我慢だ(aana)」のように名詞と助動詞の組み合わせで踏むことができるといったものだ。このように助動詞を考慮することによって、文字数の違った単語も候補に入れることができ、韻表現の幅が大きく広がる。

今回は体現に接続できる「だ」「だい」「か」「かひ?」「です」「じゃん」を名詞のあとに接続可能なものとして用いる。

6. 頭韻、脚韻

韻には「頭韻」「脚韻」という2種類がある。頭韻とは単語の最初から途中までが部分一致しているものを指し、脚韻とは単語の途中から最後までが一致するものを指す。韻は完全一致していなくてもある程度以上一致していれば十分韻として映えるものが多く、完全一致であることは求められていない。

7. 意味

ただ音が近いだけでは歌詞として不成立になるため意味が通る必要がある。今回は単語の意味的な近さを計量するのに単語の共起関係を用いる。

本稿では word2vec を用いて単語の共起関係のソートを行い、出現頻度の高いものを上位に表示するようにする。これにより数多ある単語の中からできるだけ使えそうなものをすばやく見つけることができるようになる。

8. 実験

辞書データとして mecab 辞書及び現代語を含めたライブラリである mecab-ipadic-NEologd の単語をデータベースに、「湘南台」という単語で実験をした。通常の母音の完全一致の検索結果であると以下の表 2 ように 21 件のみ韻表現として生成される。

表 2 母音完全一致の韻表現

単語	読み	母音
降誕祭	コータンサイ	o-anai
抗癌剤	コーガンザイ	o-anai
昭産開発	ショーサンカイハツ	o-anaiau
常盤開発	ジョーバンカイハツ	o-anaiau
甲南大学	コーナンダイガク	o-anaiau
	...	
大安在	オーアンザイ	o-anai
養蚕協	ヨーサンワキ	o-anai
東南町裏甲	トーナンマチウラコー	o-anaiuao-
大安在川	オーアンザイガワ	o-anaiaa
明神崎	ミョーカンザキ	o-anai
港南台	コーナンダイ	o-anai
湘南台	ショーナンダイ	o-anai
長万崎	チョーマンザキ	o-anai
湘南町屋	ショーナンマチヤ	o-anaia

対して本稿のシステムでは以下の表 3 のように 1230 件の韻表現が生成された。

表 3 韻表現形成システムで出力された韻表現

単語	読み	類似度
湘南台	ショーナンダイ	1
港南台	コーナンダイ	0.8289431
沼南	ショーナン	0.781219
徳庵	トクアン	0.7194429
岡南	コーナン	0.70537734
覚王山	カクオーザン	0.690049
陽南	ヨーナン	0.6866595
仏生山	ブツショーザン	0.6808864
城南	ジョーナン	0.66424084
港南	コーナン	0.6610156
境南	キョーナン	0.6438626
長南	チョーナン	0.64054143
六反	ロクタン	0.64048886
天保山	テンボーザン	0.628505
小田急不動産	オダキューフドー	0.61690325
	...	
海賊版	カイゾクバン	-0.2008346
増炭	ゾータン	-0.2019055
創案	ソーアン	-0.2110307
公刊	コーカン	-0.2285295

これらの中から韻が踏めそうだと感じたものを上げると「知能犯かい?」「言語道断だい」「阿鼻叫喚だい」などが挙げられる。母音の完全一致の検索結果では非常にマイナーで踏みづらい単語ばかりであったが、本稿にシステムを用いると実用的な韻表現を得ることができた。

9. おわりに

本稿では、(1)韻表現特有の音の崩し方 (2)異なる品詞の組み合わせ (3)意味的関連性 の3点に着目して、音と意味を関連性計量により適切な韻表現を形成する韻表現形成システムを示した。本システムでは、1 単語データベースを対象としたが、単語の傾向ごとに分けた複数の単語データベースを参照することによって、単語の「ジャンル」を前提とした韻を踏むことができるようになる。

今後の課題として、意味合いの計量に word2vec を使ったが、これを使うと上位に同じ種類の単語（今回の場合であると近隣の地域の名前）ばかりが出てきており、むしろ下位（もとの単語との共起性が低い）のほうが斬新さがある韻表現が多いため、意味合いの計算をさらに改良していく。

参考文献

- [1] 河本 穰, 清木 康, 藤島 清太郎, 相磯 貞和, “専門領域ドキュメント群を対象とした意味的連想

検索空間の生成方式”, 情報処理学会論文誌データベース Vol.47.

- [2] Zebra, “ジブラの日本語ラップメソッド”, 文響社, 2018.