

## 科学研究費助成事業 研究成果報告書

平成 27 年 6 月 10 日現在

機関番号：32404

研究種目：基盤研究(C)

研究期間：2012～2014

課題番号：24520173

研究課題名(和文)江戸小唄の中の舞踏的身体 - テキストマイニングによる小唄と舞踊小唄の特徴比較研究

研究課題名(英文) Dancing body in Edo Kouta - A comparative study for features of Kouta and Buyou Kouta songs using text mining methodology

研究代表者

矢島 ますみ (YAJIMA, MASUMI)

明海大学・経済学部・教授

研究者番号：80220135

交付決定額(研究期間全体)：(直接経費) 2,600,000円

研究成果の概要(和文)：本研究は「江戸小唄」と「舞踊小唄」の言語的特徴(自立語や語の連続等)を、バイグラム(2単語のペアを分析対象とするもの)を対象としたクラスター分析等のテキストマイニング手法による数理的分析を行い、小唄の“ことば世界”にみる舞踏的身体世界の独自性の解明を目的とする。「舞踊小唄」「一般小唄」の歌詞、および小唄の辞書を電子データ化し、形態素解析システム MeCab 用ユーザ辞書形式に変換しシステムに組み込んだ上で、小唄の歌詞に対して名詞・動詞・形容詞に絞った上でバイグラムを対象としたクラスター分析をおこなうと、舞踊小唄と一般小唄、および舞踊小唄の作家別のクラスタリングが出来ることが確認された。

研究成果の概要(英文)：The Object of this study is to unveil of individuality of "dancing body world" through "lyrics world" of Kouta, Japanese ballad-type song. We execute mathematical analysis using text mining techniques such as cluster analysis for lyric word bi-grams of Edo Koutas and Koutas for dancing. We digitize lyrics of all Koutas and Kouta dictionaries, and the dictionary data are introduced as additional vocabulary of MeCab, a Japanese token analysis system. We extract nouns, verbs, and adjectives from Kouta lyrics and then make up into bi-gram data. We have applied the data for cluster analysis, and realized that normal Koutas and Koutas for dancing are separated as well as Koutas for dancing themselves are separated by song writers.

研究分野：舞踊教育学

キーワード：舞踊小唄 テキストマイニング 形態素解析 クラスタ分析 バイグラム

## 1. 研究開始当初の背景

### (1) 主題となる課題と研究の動向

小唄は、安政年間、清元のお葉(1840-1901)によって創始され、幕末から明治、大正、昭和、平成に至る今日まで伝えられる江戸情緒が盛り込まれた日本独自の表現分野であるが、大衆的・庶民的な文化ゆえに学術的な研究対象になりにくかった分野である。また、文学や詩歌の世界の分析的研究は、比較的主観的な解釈や指標でなされる傾向があり、客観的に明確に語る事が難しいジャンルとも捉えられている。

しかし、近年では、“定型化されていない文章の集まりを自然言語解析の手法を使って分析して有用な情報を抽出”する「テキストマイニング手法(text mining)」が登場し、文学への応用には、竹田正幸、福田智子らの「文学作品からのテキストマイニング：文学における発見を支援する(2001)」、古典和歌からの知識発見：モビルスーツを着た国文学者(<特集>人文科学における情報処理)(2002)」などの研究成果がある。この手法は、医療や福祉、ビジネスの方面でも、言語理解とコミュニケーションとの分析手法として応用され始め、人の心・情緒を客観的に分析する言語分析方法として発展途中の手法である。

### (2) 研究対象としての小唄・舞踊小唄

今回、研究対象とする「小唄」は、大衆の感覚的な世界の表現であり、分析指標を導き出すには困難も考えられる。しかし、日本の舞踊は、バレエやモダンダンスといった“リズム”を中心としたダンス世界と異なり、“日常的な立ち居振る舞いの延長線上”にある点で、歌詞と舞踊動作との連関が強いであろうことが予測され、テキストマイニングという数理的な言語分析の手法で「小唄」と「舞踊小唄」を比較することで、粋でいなせな江戸情緒性と舞踊的身体世界(小唄振り)の独自の連関性を客観的に見つけることが出来るのではないかと考えた。

## 2. 研究の目的

### (1) 日本ならではの舞踊と小唄

小唄は、一曲2~3分という短い時間の中に、多くの意味を含ませた感覚的、大衆的な文化であり、江戸情緒を盛り込んだ三味線のテンポのよい唄である。唄に興じて踊りを即興でつけて楽しんだ時代を経て、大正時期のSPレコードの発売によって音楽は庶民の日常生活に近づき、テンポのよい小唄で踊る舞踊小唄ジャンルが大正後期から昭和初期に成立する。小唄の舞台公演や発表会では、唄に舞踊を振付けた“小唄振り”が最後に披露されて華やかに幕を閉じることが恒例である。

いわゆる「小唄」は約800曲、また、舞踊

の振付を伴った「舞踊小唄」と銘々される曲が約150曲程あると思われる。

本研究では、「舞踊小唄」の粋でいなせな江戸情緒性の特徴を、唄うことを主とした「小唄」と、舞踊の振りがついている「舞踊小唄」の両者の言語的特徴(自立語や語の連続等)を数理的分析(テキストマイニング text mining)によって、「小唄」と「舞踊小唄」のことはについて比較検討し、日本ならではの小唄の“ことば世界”と舞踊世界の連関性の存在(特徴)を捉えることを目的とする。

### (2) 研究の意義

舞踊を巡る客観的データ分析の研究傾向では、身体表現をモーションキャプチャ・システムによってデジタルデータ化し、動きそのものの特徴を捉えて比較する研究が昨今の流れである。

本研究は、舞踊や伝統芸能などの日常の延長線上にある無形文化財を、テキストマイニング等の数理的な手法を応用することで、言語文化と密接にリンクする日本の豊かな芸能諸側面の特徴を浮かびあがらせようとするものであり、具体的な舞踊的身体の客観的なデータ化を行う研究とは異なるものである。

数理的な言語分析によって舞踊小唄の特徴を導き出す今回の研究は、日本の庶民的な小唄世界と舞踊の密接な関係(日常性に潜む舞踊的身体)を改めて再発見する試みと考えている。

このような“小唄”世界の数理的な分析は、日常の延長線上の仕草や所作を含んだ日本特有の小粋な舞踊表現文化を、積極的に解明しようとする独自の研究と考えられるものであり、また、言葉で表現し難い日本芸能文化の艶や粋と言った魅力を、客観的に語る一助となる。

## 3. 研究の方法

### (1) 研究の方法(概要)

本研究では、清元のお葉(1840-1901)によって創始され、幕末から明治、大正、昭和、平成に至る今日まで伝えられる大衆芸能の小唄という世界を対象に、いわゆる唄うことを主とした“小唄”と、振付がついた“舞踊小唄”の歌詞内容の特徴を

a. 舞踊小唄と作詞家との関係

b. 舞踊小唄と一般小唄の差異表現

の視点から、テキストマイニング手法(注)によって分析し、検討を行った。

(注)『テキストマイニング』とは、定型化されていない文章の集まりを自然言語解析の手法を使って単語やフレーズに分割し、それらの出現頻度や相関関係を分析して有用な情報を抽出する手法

## (2) テキストのデータ化

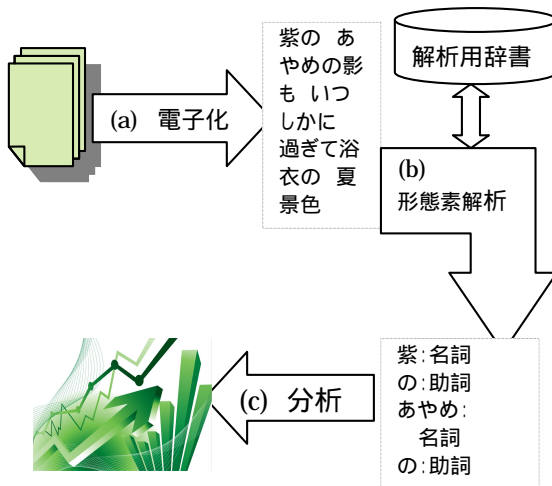
「復刻 新編「小唄」ベスト 120 (日本伝統文化振興財団 2005)」の小唄の歌詞(140 曲)と、昭和初期 S P レコード「舞踊小唄名曲集 96 曲」、および辞書類として「小唄ことばの手引(吉田 忠(編集)東京邦楽出版社 1963)」、「小唄ことば辞典(上村幸以著)」をスキャナーでスキャン後、OCR ソフトによりテキスト化を行った。しかし、完全に自動的に正しく変換されることは無かったので、アルバイト学生を集め、人手により確認および修正の作業をおこなった。

## (3) テキストマイニングツールの整備

上と並行して、テキストマイニングツールの整備を行う。分析の基本ツールとして、RMeCab(引用 )を利用した。RMeCabとは統計解析環境である R から、日本語形態素解析器 MeCab を利用するためのインターフェイスであり、形態素解析と統計解析処理を統合的に行えると考えた。

ただし、本ツールは古典語に対応していないので、「小唄ことばの手引(吉田 忠(編集))」、「小唄ことば辞典(上村幸以著)」を電子化して、上の辞書へ移植・統合して、さらに解析精度をあげた。

研究の流れは下図の通りである。



## 4. 研究成果

### (1) 西條八十と平山蘆江の舞踊小唄の比較

「舞踊小唄名曲集 第1集~第4集(96 曲中)」において、西條八十(さいじょうやそ)と平山蘆江(ひらやまろう)が作詞した小唄の歌詞をそれぞれ 35 曲、34 曲分用意し RMeCab による分析を試みた。RMeCab とは統計解析環境である R から、日本語形態素解析器 MeCab を利用するためのインターフェイスであり、形態素解析と統計解析処理を統合的に行えるテキストマイニングツールである。

本分析においては、それぞれの作詞家の曲を 3 つにわけ、それぞれ saijo\_1.txt, saijo\_2.txt, saijo\_3.txt, hirayama\_1.txt, hirayama\_2.txt, hirayama\_3.txt という文書ファイルを作成し、RMeCab により名詞、動詞、形容詞を抽出した。また、データ間距離についてはキャンベラ距離を導入した。キャンベラ距離は、次式で定義されるものである。

$$\sum_{i=1}^n \left| \frac{x_i - y_i}{x_i + y_i} \right|$$

RMeCab を用いて分析を行った結果を図 1 に示す。西條、平山のそれぞれの作家の 3 ファイルがクラスターを構成し、2 作家の歌詞が自動的に分類されることが確認できる。

つぎに、潜在的意味インデキシングによる分析について述べる。潜在的意味インデキシングは、ターム・文書行列を圧縮することで、分類を効果的に行う技術である。高次元のターム・文書行列を低次元で近似的に表現する技術とも言える。これは、もとのターム・文書行列に特異値分解を行うことで実現する。本分析では、クラスター分析で用意したデータファイルを再度用いて、それらに対して特異値分解を行った。その結果を表示したものを、図 2 に示す。この図においても hirayama と saijo のそれぞれの文書が近傍に集まっていることが明らかであり、これはすなわち作家による歌詞の分類が可能であることを表している。

以上により、テキストマイニング技術による自動分類の可能性が認められた。

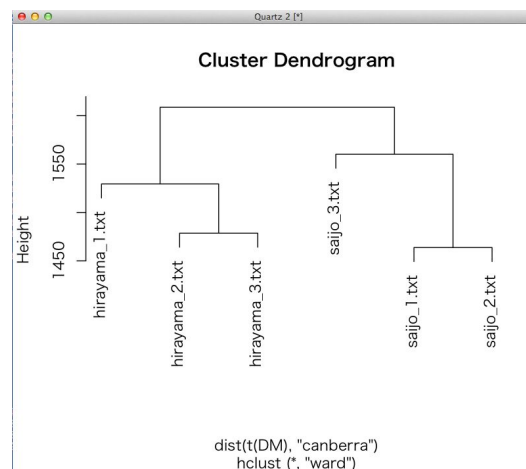


図 1: クラスター分析出力

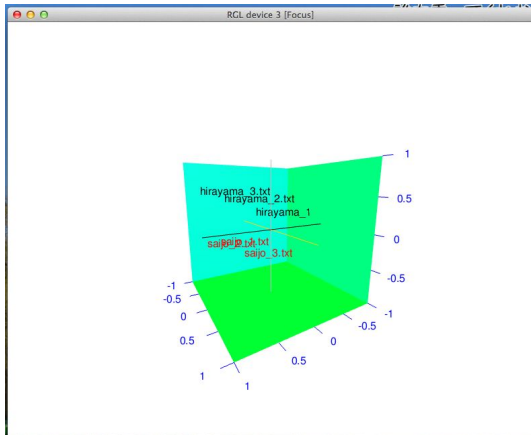


図2：潜在的意味インデキシング出力

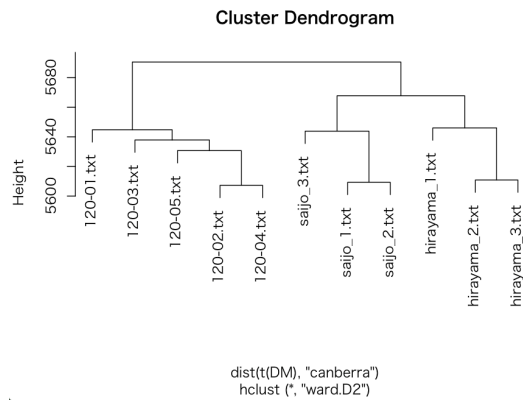


図3：クラスター分析出力（一般小唄含む）

表1 「舞踊小唄名曲集（96 曲中）」で西條八十と平山蘆江が作詞した曲

西條八十	お染、お富、俳聖芭蕉、仏蘭西人形、柳二題、狐島田、正行、忘れな草、東京三番叟、蛸をどり、新唐人お吉、沖のかもめ、梅暦、春雨獅子、嶋の朝比奈、乙女椿、霊峰富士、紀文、金魚、白衣の乙女、夕顔の巻、須磨の巻、浦島、孝女白菊、お七人形振り、紅日傘、舞踏会への招待、あやめ娘、芽ふき柳、せせらぎ、江戸むらさき、利根の夜船、夏木立、手兒奈、もらひ水
平山蘆江	鐘馗、身代わり座禅、神楽面、兩國夜景、玉菊灯籠、寿三番、扇かざして、藤袴、文楽、土橋の雨、紅小袖、おもかげ、水鏡、浮名小唄、かすみ、松葉かんざし、河童、花笠、凱旋、伴内、つく羽根、お初、とろろ、今様手習子、丹次郎、村そだち、男鬚、對の網笠、雨やどり、孫悟空、新鳥羽絵、日の丸、能因法師、かるた会

(2)舞踊小唄と一般小唄の差異表現の抽出

小唄ことばの解説書である「小唄ことばの手引」「小唄ことば辞典」を電子化したデータを、形態素解析システム MeCab 用ユーザ辞書形式に変換しシステムに組み込んだ上で、再度分析を行った。その結果、名詞・動詞・形容詞に絞った上でバイグラムを対象としたクラスター分析をおこない、舞踊小唄と一般小唄のクラスタリングが出来ることが確認された。さらに舞踊小唄についても、西條・平山の作詞家別のクラスタリングも出来ており、本研究の重要な成果となった（図3参照）。

具体的な特徴を抽出していくことが今後の課題である。

< 引用文献 >

石田基広：「Rによるテキストマイニング入門」、森北出版、2008。

5．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕(計 2 件)

矢島ますみ、前田利之、"舞踊小唄の特徴パターンの分析 - テキストマイニング技術を用いた歌詞による傾向の抽出・分類の検証 -"、明海大学教養論文集「自然と文化」、査読有、No.24、2013、39-43

前田利之、データ評価のためのデータマイニング技術の活用、電気学会研究会資料集、IIC-14-159、2014、115-119

6．研究組織

(1)研究代表者

矢島ますみ (YAJIMA, Masumi)  
明海大学・経済学部・教授  
研究者番号：80220135

(2)研究分担者

前田利之 (MAEDA, Toshiyuki)  
阪南大学・経営情報学部・教授  
研究者番号：70320041