

第23回公開シンポジウム  
人文科学とデータベース  
発表論文集

Proceedings of the 23rd Symposium  
on Humanities and Database

2018年3月3日(土)  
大阪府立大学  
B3棟116 中会議室

- 【主催】 第23回公開シンポジウム「人文科学とデータベース」実行委員会  
【共催】 人文系データベース協議会・大阪府立大学 情報システム研究センター  
【後援】 アート・ドキュメンテーション学会関西地区部会  
情報処理学会人文科学とコンピュータ研究会  
情報知識学会関西部会

## 目 次

### ●招待講演

- 中尾佐助探検記録（タイプスライド）のデータベース化について  
山口裕文（大阪府立大学名誉教授）……………1

### ●一般講演 1

- 「西浦の田楽」で継承される「教え」の映像デジタルアーカイブの構想  
彦坂和里（目白大学），杉山岳弘（静岡大学）……………3
- 浜松市における無形民俗文化財「祭り・神事・行事」のデータベース化と継承状況の調査  
杉山岳弘（静岡大学），戸田剛，太田好治（浜松市）……………9
- 語の出現傾向による『源氏物語』第一部各巻の分類  
土山玄（一橋大学）……………17

### ●一般講演 2

- 中尾佐助スライドデータベースと大学リポジトリとの連携に向けて  
小島篤博，青木茂樹，泉正夫，宮本貴朗（大阪府立大学）……………23
- 郷土食による地域理解支援システム「もちマップ」のオープンデータ化についての考察  
河村郁江（名古屋工業大学），伊藤宗太（インフォ・ラウンジ合同会社），  
伊藤孝行，白松俊（名古屋工業大学）……………29
- 『本草和名』データベースの構築に向けて  
武倩，劉冠偉（北海道大学）……………37

### ●一般講演 3

- 方言分布・言語地図データベース —時空間情報を持つ言語データ—  
大西拓一郎（国立国語研究所）……………43
- 上智大学「日本の人名データベース」の構築について  
ボルン礼於（ハイデルベルグ大学）……………51

## 招待講演

# 中尾佐助探検記録(タイプスライド)のデータベース化について

山口 裕文

(大阪府立大学名誉教授)

自然や人間生活の実地調査では得た情報の根拠を残すのが難しく、多くの場合、紀行文の記録だけになって再確認できないため、実験科学的研究者からの評価を受けられない事がある。フィールド研究では再現性を確認できる採集物や種子などで根拠を確保し、生物学での新種の記録(記載 **description**)では根拠のタイプ標本を指定する。タイプは実物標本でも描画でも認定されるが、タイプの無い記述では正式な種名には認定されない(裸名という)。中尾佐助は探検で撮影したスライドをタイプスライドと呼んでフィールド調査の根拠としている。押し花の標本と違って写真画像は永久に保存できず劣化するので、解決の手法としてデジタル化した画像をデータベースに構築したのが府立大学図書館で公開している中尾スライド DB である。本講演ではいくつかの探検を例としてフィールドノートや著作との連携を紹介する。

「西浦の田楽」で継承される「教え」の  
映像デジタルアーカイブの構想  
**The Initiative of the Video-based Digital Archive  
for Conserving the “Oshie” in “Nishiure Dengaku”**

彦坂 和里<sup>\*1</sup> 杉山 岳弘<sup>\*2</sup>

Airi HIKOSAKA<sup>\*1</sup> Takahiro SUGIYAMA<sup>\*2</sup>

目白大学 社会学部, 東京都新宿区中落合 4-31-1<sup>\*1</sup>

静岡大学 情報学部, 静岡県浜松市中区城北 3-5-1<sup>\*2</sup>

Mejiro University, 4-31-1, Nakaochiai, Shinjuku, Tokyo<sup>\*1</sup>

Shizuoka University, 3-5-1, Johoku, Nakaku, Hamamatsu, Shizuoka<sup>\*2</sup>

**概要:** 本研究では、国指定重要無形民俗文化財「西浦の田楽」を保存し継承を支援するため、これまでに全演目の動画撮影および簡易アーカイブの構築を試験的に実施してきた。本研究の最終目標は、別当(祭主)および能衆(舞手等の役割を担う者)が継承している「教え」(舞い方や信仰心なども含め、先代から言われ継承される知識全般)を、筆者がこれまでに作成した形の知識構造の枠組みを用いて体系的にモデル化し、映像デジタルアーカイブ化することにある。本稿では、これまでの試みおよびアーカイブ化の構想について述べる。

**Abstract:** "Nishiure Dengaku" is an important intangible folk cultural asset. To conserve "Nishiure Dengaku" and help with succession, we have tried to take the videos of all performances and build a text-based database. The objective of this study is to establish the video-based digital archive of the “Oshie” which is the knowledge included devotion inherited from the predecessor. To realize this, we will modelize the “Oshie” based on the framework of knowledge structure of Kata which created in our earlier study. In this paper, we describe the approach and initiative of the video-based digital archive.

**キーワード:** 西浦の田楽, 民俗芸能, デジタルアーカイブ, VR コンテンツ

**Keywords:** Nishiure Dengaku, folk performing arts, digital archive, VR contents

## 1. はじめに

本研究は国指定重要無形民俗文化財「西浦の田楽」を研究対象とする。「西浦の田楽」とは、静岡県浜松市天竜区水窪町にある西浦観音堂で約 1300 年もの間継承されている民俗芸能である。西浦の田楽の始まりは諸説あるが、養老 3 年(719 年)に行基菩薩が西浦の地を訪れ、仏像を観音堂に安置して以降、行われるようになったとされている[1]。

毎年旧暦 1 月 18 日の月の出から翌 19 日の日の出にかけて、全 47 演目(旧暦で閏年にあたらぬ年は 46 演目)を奉納する。この田楽では、別当および能衆を担う各家での口伝による世襲により、日本中世の舞

の姿をほぼ変えずに継承している。その文化的価値は非常に貴重で、今後も継承していくべき地域文化である。しかし、地域の過疎化や少子高齢化による後継者不足のため失伝の危機にある。

そこで、本研究では、西浦の田楽で継承される知識を保存し、継承を支援することを目的とし、映像デジタルアーカイブを構築する。

本稿では、西浦の田楽とその継承に関して述べた後、本研究の構想と、これまでの先行研究および試験的に実施してきたテキストベースの簡易アーカイブ化、構想の実現に向けて行った全演目の動画撮影について述べる。

## 2. 西浦の田楽の概要

西浦の田楽で奉納される演目とその継承について述べる。

### 2.1 西浦の田楽の演目とその継承状況

西浦の田楽は全部で 47 演目あり、2 種類に大別される(表 1,2)。「地能」は 33 演目が該当し、継承する能衆や奉納順が決まっている。ただし、病気等のやむを得ない事情でその能衆が舞えない場合は、別の能衆が担当することもある。「はね能」は 12 演目が該当し、担当する能衆が決まっておらず、年によって舞い手や奉納順が異なる。このうち「閏舞」は旧暦で閏年にあたる年のみ奉納される。残りの 2 演目は「番外」とされ、「タイハイ」と「しづめ」が該当する。

表 1 地能 33 演目

No.	演目名	No.	演目名
1	庭ならし	18	よなぞう
2	御子舞	19	鳥追い
3	地固め	20	殿舞
4	地固めもどき	21	早乙女
5	劔	22	山家早乙女
6	劔もどき	23	種おり
7	高足	24	桑とり
8	高足もどき	25	糸引き
9	猿の舞	26	餅つき
10	ほた引	27	君の舞
11	舟渡し	28	田楽舞
12	鶴の舞	29	佛の舞
13	出体童子	30	治部の手
14	麦つき	31	のた様
15	田打ち	32	翁
16	水な口	33	三番叟
17	種まき		

表 2 はね能 12 演目(平成 28 年度の奉納順)

No.	演目名	No.	演目名
1	高砂	7	屋島
2	しんたい	8	山姥
3	梅花	9	さお姫
4	惶々	10	野々宮
5	観音の御法楽	11	閏舞
6	鞍馬天狗	12	辨慶

別当・能衆を担う家は決まっているが、近年では後継者不足の問題から、もともと能衆の家系ではない者が能衆に加わることもある。現在能衆は 17 名で構成されている。

西浦の田楽では、別当・能衆が持つ知識は短期間に次の代へ渡されるのではなく、生涯にわたって徐々に渡されていく。例えば、幼少期の頃から親や地域住民が日常的に西浦の田楽について子に話したり、舞を真似させたりする。子が能衆として経験を積み、親が子を能衆として認めるとき等に世代交代が生じる。世代交代の時期に決まりはなく、各家での判断に任されている。継承方法は口伝であるが、家によってはその家に伝わる資料や舞う姿を記録した動画等を通して学習することもある。

### 2.2 西浦の田楽の継承に関する本研究の考え方

本研究では、別当や能衆が世代間で継承する知識全般を「教え」と定義する。「教え」には、同じ演目の中で繰り返し舞う回数、舞う方角順、その家が担う役割等が含まれると考える。

この「教え」の継承に関する本研究の仮説として、「先代から受け継いだ「教え」に、個人の経験や考えを反映させた状態で、次の代へと継承している」と考えている(図 1)。

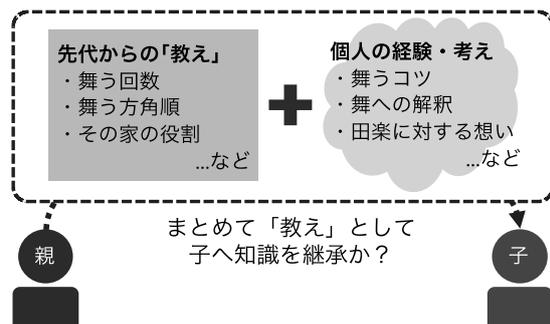


図 1 「教え」の継承の仮説イメージ

筆者が参加した西浦の田楽保存会主催の平成 29 年度の学習会(平成 29 年 6 月 18 日開催)にて、能衆らは「舞う回数や方角の順番があり、それは必ず守らなくてはならない」旨を参加者へ繰り返し説明していた。これは代を経ても変わらず継承される知識である。その一方で、個人が経験によって得た舞うコツや舞に対する解釈、西浦観音堂への信仰心等の考え方は個人によって異なるほか、同じ個人でも年を重ねるごとに考え方が変わっていくこともある。本研究では、先代からの「教え」に対し、能衆らは各自の経験や考え方に基づく知識を反映させて、まとめて次の代へ「教え」を継承していると考えている。

そのため、映像デジタルアーカイブを構築する際に、同じ家や個人に関する記録を取るにしても、記録する年によって舞に対する考え方が変化する可能性があると考えられる。つまり、西浦の田楽の「教え」のアーカイブ化は、年を経るとともに変化していく別当・能衆という人物のアーカイブ化を目指すこととなる。このた

め、ある1年の記録を残すだけでは、「教え」のアーカイブと呼ぶには不十分であり、変化していく個人の知識、時系列なデータを継続的に保存し、参照できるアーカイブの仕組みを検討する必要がある。

### 3. 本研究で目指す映像デジタルアーカイブの構想

西浦の田楽の継承状況および本研究の仮説を踏まえ、本研究で構築する「教え」の映像デジタルアーカイブの構想について述べる。デジタルアーカイブのイメージを図2に示す。

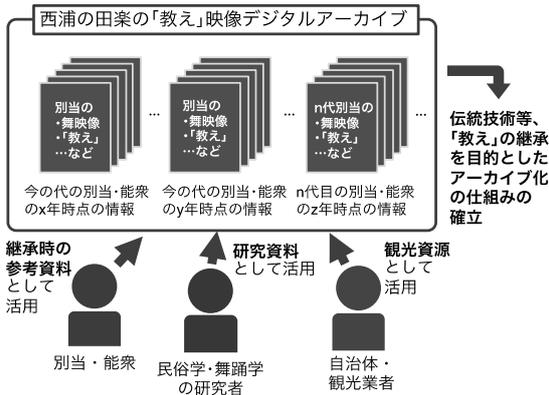


図2 「教え」の映像デジタルアーカイブのイメージ

単に当代の別当・能衆のある年の情報をアーカイブ化するだけではなく、同じ別当・能衆の別の年の時点での情報や別の代の情報も記録・参照できるようにする。

これにより、別当や能衆にとっては継承時の参考資料とすることができる。また、民俗学や舞踊学の研究者には研究資料になるほか、自治体や観光業者がアーカイブ内の情報を観光資源として活用できると考える。さらに、このアーカイブ化の仕組みは民俗芸能だけでなく、先代から次の代へと継承を行う伝統技術等の分野のアーカイブ化の仕組みを確立することにつながる。

これを実現するためには、世代を越え変わりゆく情報をどう永続的に記録するか、変わりゆく情報を正確に記述できるデータ構造とは何か、変わりゆく情報を検索できる仕組みとは何か、「教え」をコンテンツとしてどのように表現するか、といった課題に対する解決策を検討していく必要がある。

これらの課題に対する解決策の検討に関連してこれまでの研究と予備調査について述べる。

#### 3.1 形の知識構造の枠組みの作成

これまで筆者は剣道における形(かた)である「日本剣道形」に着目し、形教育の研究を進めてきた[2-5]。剣道形を授業で体系的に教育する環境を構築する過

程で、形の知識構造の枠組みを作成した。この枠組みを用いて剣道形の知識を体系的に記述した[5]。

形の知識構造の枠組みを図3に示す。形の知識構造を以下の4層に分けている。

層	定義	該当する知識	西浦の田楽での例
心的知識	形で示されている教えに関する知識	形に示される教え(心)	・豊作祈願 ・子孫繁栄
意義的知識	所作の背景に潜む考え方に関する知識	所作にある考え方、由来	・早乙女の由来 ・なぜ必ず五方を踏むのか
論理的知識	所作を行う意味に関する知識	・用語としての意味 ・所作を行う意味	・五方を踏む意味 ・五方とは5つの方角のこと
身体的知識	実際の身体動作に関する知識	・所作のポイント(特徴) ・一連の流れ(流れ)	・必ず五方を踏む ・早乙女の流れ

図3 形の知識構造の枠組み

#### (1) 「身体的知識」の層

この層は実際の身体動作に関する知識の層である。所作の流れや各所作のポイントは、いずれも頭で理解するだけでなく、実際に体を動かしてこそ習得できる身体的な知識である。この層では、所作の流れは「流れ」、所作のポイントはその所作を構成する特徴であると考え、「特徴」として記述する。西浦の田楽で言えば、「早乙女」の流れ、必ず五方(5つの方角、中央・東・西・南・北のこと)を踏むこと等が該当すると考える。

#### (2) 「論理的知識」の層

この層は所作を行う意味に関する知識の層である。所作を行う意味は、形の所作と精神性の関連性を論理的に説明する上で重要な知識にあたる。西浦の田楽で言えば、五方を踏む意味、五方という言葉の意味等が該当すると考える。

#### (3) 「意義的知識」の層

この層は所作の背景に潜む考え方に関する知識の層である。所作の背景に潜む考え方は、「論理的知識」の層と同様に形の所作と精神性の関連性を論理的に説明し、その所作を正しく行う上で、重要な知識にあたる。西浦の田楽で言えば、「早乙女」の由来、なぜ必ず五方を踏まなければならないのか等が該当すると考える。

#### (4) 「心的知識」の層

この層は各形で示されている教え・教義に関する知識の層である。一つの形を通してどのような先人たちの教え・教義が表現されているかを記述する。西浦の田楽で言えば、豊作祈願や子孫繁栄への祈り等が該当すると考える。

西浦の田楽は所作と精神が切り離せない点で剣道形と類似しており、同じく形の知識構造の枠組みを適用して「教え」を体系的に記述できる可能性がある。ただし、西浦の田楽は地域の五穀豊穡や水難・火難除け等を願って行われる神事であり、心的知識の

層よりもさらに深い宗教的な知識が内在している可能性があると考える。「教え」を現在の形の知識構造の枠組みだけでは表現しきれない場合、枠組みを基にして新たな記述方法を検討する。

### 3.2 演目の簡易アーカイブの試験的な作成

「教え」を収集し形の知識構造の枠組みを用いて体系的に記述するにあたり、文献等ですでに明らかになっている演目に関する知識を事前に整理するため、テキストベースの簡易アーカイブを試験的に作成した。

今回は『西浦田楽の民俗文化論』[1]と菅原ら(2005)[6]を用いた文献調査、および後述の撮影動画、筆者が参加した平成 28,29 年度の西浦の田楽学習会から情報収集を行った。全 47 演目について収集した情報を、種類分けして整理し、表計算ソフト上にまとめ、全 13 カラムのテキストベースの簡易アーカイブを作成した。表 3 に地能「早乙女」の例を示す。

項目名について説明する。「番号」は演目に割り振った固有の番号である。なお、お囃子の項目に用いている言語的表現は文献[1]に書かれているものを主に採用し、文献[1]にお囃子の記載がなかった演目については撮影動画を確認し、□・○・△へ記号化してお囃子の種類を分けて入力した。登場する面についても、同様に文献[1]に記載のないものは撮影動画を確認し入力した。

簡易データベースを作成した結果、地能「治部の手」からはね能全 12 演目にかけて、同じお囃子が使われていることを可視化できた。また、「御子舞」、「猿の舞」、「田楽舞」のお囃子は、その各演目でしか演奏されないものであることも可視化できた。今後は整理した情報をもとに「教え」の収集を行う予定である。

## 4. 全演目の試験的な動画撮影

次に、デジタルアーカイブ化の構想の実現に向けて、平成 28 年度の西浦の田楽開催時(平成 29 年 2 月 14 日(火)～15 日(水))に行った、全演目の試験的な動画撮影について述べる。

### (1) 使用機材

今回使用した機材の構成は次のとおりである。

- ・デジタルビデオカメラ 8 台(Canon HF G20)
- ・カメラバッテリー 25 本(Canon BP-819D\*17 本、BP-808D\*8 本)
- ・三脚 2 本(Velbon ULTTREK UT60、SLIK U9800)
- ・ワイドコンバージョンレンズ 1 本(Canon WD-H58W)

### (2) 会場の概要

演目を奉納する会場内の配置とおおよそのカメラの設置位置は図 4 のとおりである。図中の「ガクトウ」はお囃子や謡を担当する能衆が座する場所である。「幕屋」

表 3 演目の簡易アーカイブの例「早乙女」

項目名	データ
番号	24
名称(よみがな)	早乙女(そうとめ)
種別	地能
内容	この次第は、前半の「花ざさら」と後半の「はんこいつき」の二部構成である。(中略)前半の「花ざさら」は次のように舞う。觥を左肩に担いだ二人が出て、太鼓と「チャーチャーチーチイ、チララウラーラ」の笛、そしてその口拍子で、二人向かい合って位置交代をしながら觥をすって舞う。(以下、紙面の都合上省略)
舞手の人数	3(前半 2 人、後半 1 人)
時間の長さ(分)	10
表現しているもの	田植え、稲作の耕作過程
全体の意味	神への祈り
祈り	豊作祈願、子孫繁栄
舞の種類	農耕、稲作と養蚕の田遊
囃子・口拍子の種類	前半「チャーチャーチーチイ、チララウラーラ」、後半「ネーギネーギ、ナンダラヨ」
登場する面	なし
備考	並行して、「ネンネン坊のご飯(白米)」を喜平地が観音・鎮守・愛宕に供える。これは「殿舞」が始まると別当屋に下りてもらってくる。しかし、稗団子も「ネンネン坊のご飯」という。

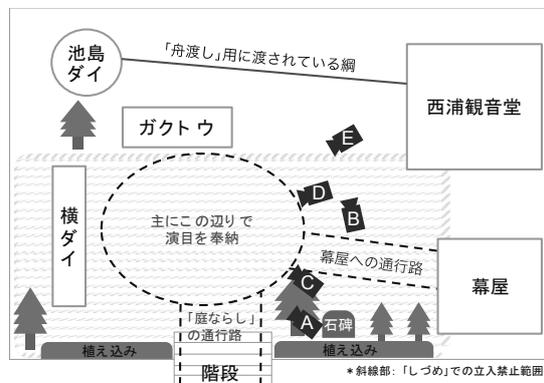


図 4 会場内の配置と撮影位置の略図

は能衆らが出番以外のときに待機する小屋である。「池島ダイ」は「舟渡し」の際に灯される、高さ 13 尺(約 4m 強)の柱状の大きな松明である。「横ダイ」は横に寝かされた状態で積み上げられた松明である。この松明は田楽のスタート時から灯される。基本、ガクトウの前あたりで演目が奉納されるが、演目によっては、幕屋から始まるもの(例:「佛の舞」)や、幕屋や観音堂付近まで出るもの(例:「よなぞう」、「田楽舞」)、階段下から始まるもの(「庭ならし」)等もある。

#### (3) 撮影スケジュールおよび人員構成

主な撮影スケジュールは表 4 のとおりである。西浦の田楽は地域住民やアマチュアカメラマン等も見学に来るため、14 日の 16 時 30 分頃に会場へ到着し、17 時頃に図 4 の A の位置に三脚を設置し場所取りを行った。20 時 54 分頃に最初の演目「庭ならし」が始まり、翌日 8 時 35 分頃に最後の演目「しづめ」が終了し、9 時頃には撮影の撤収が完了した。

その間、筆者を含む撮影スタッフ 2 名、撮影補助の大学生 2 名で、基本 1 台のビデオカメラを使って撮影した。筆者が西浦の田楽において特徴的なものだと考えた演目(「高足」、「舟渡し」、「出体童子」、「田楽舞」)、および面をつけて行う演目のうち「猿の舞」、「佛の舞」については 2 台で撮影した。なお、前述の機材構成においてビデオカメラが 8 台となっているのは、使用したビデオカメラが内蔵 HDD に記録するものであり、残量が少なくなった際にカメラを本体ごと交換するようにしたためである。また、寒さによりバッテリーの持続時間低下、カメラ本体の機能低下のおそれがあったため、適宜使い捨てカイロを使って機材を温めながら撮影を行った。

#### (4) 撮影位置

西浦の田楽では奉納場所の範囲を明確に示すことなく行われる(ただし、後述の「しづめ」を除く)ため、カメラの前に他の見学者が来ることがあった。そのため、撮影位置は演目によって移動している。カメラの移動タイミングについては前述の表 4 の「撮影位置」列に記載のとおりである。表中の A~E は図 4 中の A~E と対応している。

「舟渡し」は西浦観音堂から池島ダイに架けられている綱を使い、舟の形をした松明を渡して池島ダイを灯す演目である。そのため、他の演目と奉納場所が異なることから、C の位置で撮影した。最後に行われる「しづめ」では、奉納直前に結界を示すしめ縄が張られ、図中の斜線部分が立ち入り禁止エリアとなる。そのため、図中の E の位置で撮影した。

#### (5) 撮影時の課題点

撮影の結果、録画データの総時間数は約 8 時間半

表 4 主な撮影スケジュール

時刻	内容	撮影位置
14 日 13 時	静岡大学にて機材準備	-
14 時	静岡大学出発	-
16 時 30 分頃	西浦観音堂到着、下見や関係者への挨拶	-
17 時頃	撮影場所決定、場所取り	A
20 時 54 分頃	西浦の田楽開始	A
15 日 0 時 16 分頃	地能「舟渡し」	B
0 時 37 分頃	地能「鶴の舞」	C
2 時 15 分頃	くらいれ(能衆らが別当家で食事をとる)のため撮影を中断し休憩	-
3 時 36 分頃	演目の奉納再開(地能「種おり」)	C
6 時 58 分頃	はね能「山姥」	D
8 時 16 分頃	番外「しづめ」	E
8 時 35 分頃	西浦の田楽終了、能衆らの集合 写真撮影、機材撤収	-
9 時頃	撤収完了	-

であった。今回の撮影によりわかった課題点について述べる。

#### 【撮影位置】

撮影スタッフの荷物の置き場所等を考慮し、今回開始時は A の位置で撮影したが、一部の演目にて、カメラの前に他の見学者が立ってしまい、部分的にしか映せなかった。横ダイ付近は比較的に見学者が少なく、また人の移動も少ないため、撮影位置として適していると考えられる。ただし、「佛の舞」等においては幕屋から始まるほか、前述のとおり「舟渡し」は観音堂付近で行われることから、少なくとも 2 台のカメラおよび撮影人員を配置できると望ましい。また、ガクトウ前や幕屋付近等で舞う能衆まで全体をより正確に撮影するためには、全方位を撮影する必要がある。

#### 【撮影環境(寒さ、暗さ)】

西浦の田楽が行われる浜松市天竜区水窪町は長野県との県境に位置する山間の町である。そのため、西浦の田楽が開催される時期は日が落ちると気温零度を下回る。その極寒環境下で夜通し奉納するため、こまめに休憩が取れる程度の人員が必要である。

また、西浦の田楽保存会が表現する言葉を借りれば「西浦の田楽はもてなさない祭り」であり、夜間の光源は池島ダイや横ダイの火と、ガクトウや幕屋を灯す明かりのみで、見学者のために照明をつけることはない。そのため、撮影機材の選定にあたっては暗さに強いものを選ぶ必要がある。

#### 【長時間撮影可能な機材構成】

西浦の田楽で奉納される演目の長さは短いものは1分程度で終了するが、地能「御子舞」は3人の舞い手が一人約20分、計60分程度舞い続ける。今回のビデオカメラは十分耐えうるものであったが、このような長い演目でも連続して録画でき、バッテリーがもつ機材を選ぶ必要がある。

## 5. 今後の計画

これまでに行った先行研究および予備調査を踏まえ、「教え」の映像デジタルアーカイブ構築の実現に向けた今後の計画について述べる。

まず①別当・能衆へ「教え」について調査を行い、形の知識構造の枠組みをもとに収集した知識を体系的に記述する。また、②西浦の田楽の全演目をより正確に記録し、多視点から観察できるよう、VR映像コンテンツを制作する。そして、③「教え」を正確に表現できるデータ構造を設計し、デジタルアーカイブを構築する。

①では、現在の別当・能衆への聞き取り調査と継承の様子の観察、文献調査を実施し、継承される「教え」を収集する。収集した「教え」を、形の知識構造の枠組みに当てはめ、体系的に記述する。枠組みでは表現しきれない場合、枠組みをもとに新たな記述方法を検討する。

②では、西浦の田楽開催時に複数台の全方位ビデオカメラを用いて全47演目を撮影し、VRコンテンツ化する。これは当日の雰囲気を含め各演目の全体像を正確に記録とともに、多視点から観察できるようにするためである。

③では、①での体系化の結果を踏まえ、Dublin Core[7]や共通語彙基盤[8]をベースに、「教え」を正確に記述できるデータ構造を設計する。また、年ごとの記録が可能なデータベースを設計する。そして、VR映像コンテンツを使った、PCやタブレット端末等で閲覧可能なデジタルアーカイブを構築する。詳細については、「教え」を収集する過程で検討を進める。

## 6. まとめ

本研究では、国指定重要無形民俗文化財「西浦の田楽」を保存し継承を支援するため、別当および能衆

が継承している「教え」の映像デジタルアーカイブを構築することを目指している。本稿では、映像デジタルアーカイブの構想、およびこれまでの研究や予備調査、構想の実現に向けた全演目の動画撮影について述べた。今後、能衆らへの調査や観察を通して「教え」を収集し、アーカイブ化の手法を検討していく。

## 謝辞

平成28年度西浦の田楽の撮影にあたり、ご協力いただきました西浦の田楽保存会、撮影スタッフ・補助の皆様にご感謝の意を表す。

## 参考文献

- [1] 吉川裕子, 西浦田楽の民俗文化論, 岩田書院, 2012.
- [2] 彦坂和里, 杉山岳弘, 白井靖人, 杉山融, "中学校の武道教育における日本剣道形の指導支援のための電子教材の評価", 日本教育工学会第29回全国大会講演論文集, pp.483-484, 2013.
- [3] 彦坂和里, 西尾典洋, 杉山岳弘, 白井靖人, 杉山融, "日本剣道形の指導支援のための電子教材による映像提示の効果", 教育システム情報学会研究報告, vol.29, no.4, pp.31-34, 2014.
- [4] 彦坂和里, 西尾典洋, 杉山岳弘, 白井靖人, 杉山融, "日本剣道形の指導支援のための電子指導書の開発", 静岡大学情報学研究, vol.21, pp.43-55, 2016.
- [5] 彦坂和里, "日本剣道形による武道教育のためのブレンド型教育環境の提案", 静岡大学大学院情報学研究科修士論文, 2016.
- [6] 菅原和孝, 藤田隆則, 細馬宏通, "民俗芸能の継承における身体資源の再配分: 西浦田楽からの試論", 文化人類学, 70巻, 2号, pp.182-203, 2005.
- [7] 杉本重雄, "Dublin Coreについて 第1回 概要", 情報管理, Vol.45, No.4, pp.241-254, 2002.
- [8] 独立行政法人情報処理推進機構, "共通語彙基盤 | 共通語彙基盤整備事業", <http://goikiban.ipa.go.jp/>, (参照 2018-02-09).

## 浜松市における無形民俗文化財「祭り・神事・行事」の データベース化と継承状況の調査

### A Database of Intangible Folk Cultural Properties "Festivals, Shrines and Events" in Hamamatsu City and a Survey of Succession Situation

杉山岳弘<sup>1</sup>, 戸田剛<sup>2</sup>, 太田好治<sup>3</sup>

Takahiro Sugiyama<sup>1</sup>, Tsuyoshi Toda<sup>2</sup>, Yoshiharu Ohta<sup>3</sup>

1) 静岡大学情報学部, 静岡県浜松市中区城北 3-5-1

2,3) 浜松市市民部文化財課, 浜松市中区元城町 103 番地の 2

1) Shizuoka University, 3-5-1, Johoku, Nakaku, Hamamatsu, Shizuoka

2,3) Hamamatsu city hall, 103-2, Motoshiro-cho, Nakaku, Hamamatsu, Shizuoka

**概要:** 浜松市には、国・県・市に指定された無形民俗文化財を含め、古くから伝わる「祭り・神事・行事」が 300 件以上ある。研究室では浜松市と協同で、2014 年から祭りについて情報収集を実施し、その成果として、2017 年にお祭りの Web データベースを構築して公開した。また、情報収集時に、主催者に対して継承状況の調査を実施した。本稿では、無形民俗文化財のデータベース化の過程と継承状況の調査結果について報告する。

**Abstract:** In Hamamatsu City, there are more than 300 traditional festivals and shrine rituals, including the intangible folklore cultural asset specified by nation, prefecture and city. Our laboratory has worked in collaboration with Hamamatsu City and gathered the information of festivals since 2014. As a result, we released the web database of 2017's festivals that we constructed. Besides, when gathering the information, we conducted the survey of succession condition against the organizer. In this paper, we will report the survey results which is about the database establishment process of the intangible folklore cultural asset and the succession condition.

**キーワード:** 浜松市、無形民俗文化財、データベース、継承

**Keywords:** Hamamatsu-city, intangible folk cultural properties, database, succession situation

#### 1. はじめに

浜松市では、「祭り」としての行事が、365 日を通じて 300 件以上も開催されている。ここまでの祭りの伝統が市民により継承されてきたことは、誇るべきことであり、今後も継承していくべき貴重な地域文化である。この重要さは、浜松市の条例である「浜松市民俗芸能の継承及び振興に関する条例」(平成 28 年 3 月 24 日制定)に現れており、市の条例として、民俗芸能を継承・振興していくことを定めている[1]。

これらの祭りは時代とともに変化し、行事が簡略・省略されたり、場合によっては、継承が困難で消滅したり、消滅の危機にあつたりする祭りもある。まずは、現状を調査・把握し、継承に必要となる祭りの情報を収集して記録していくことが急務である。

きっかけは、浜松市市民部文化財課からの提案で、平成 26 年度の静岡大学情報学部の地域連携の取り組みの一環で「浜松おまつり暦」(Web カレンダー)を

制作[2]してからはじまる。これがもととなり、平成 27 年度に静岡大学イノベーション社会連携推進機構「平成 27 年度地域連携応援プロジェクト」に採択、平成 28 年度・平成 29 年度に浜松市「みんなのはままつ創造プロジェクト」に採択され、継続して浜松市の「祭り」に関する調査を実施してきている。特に、平成 28 年度の調査では、浜松地域の全点調査を目標に、浜松市市民部文化財課と連携し、各所の地域協働センターと協力体制を作り、祭りの基本情報の(改めての)収集と継承状況の調査を実施した。この成果をもとに、「浜松お祭りアーカイブ」を公開した[3]。

本報告は、平成 26 年度から平成 29 年度にかけて行った、浜松市におけるお祭りのデータベースの構築と継承状況に関する調査についての報告である。なお、以後、祭り・神事・伝統行事について、特に指定がない場合は、祭り、で表すものとする。

## 2. 浜松市における伝統行事などの現状

約 20 年前の平成 17 年、浜松市で大規模な合併が行われた(表1)。その際に浜松市内全域に存在する祭りや行事を把握するため、市が各自治体に対し調査を行った。結果、当時、286 件の祭りや行事が確認された。表 2 は、平成 28 年度の調査をする段階で確認されていた各区毎の祭りや行事の数である。特に中山間地域や山間部においては、古くから伝統行事が継承されてきており 200 件以上もの祭りが存在している。

表1. 浜松市における平成の合併状況

旧市町村名	合併後の区割り	合併前の所属	備考
浜松市	中区,北区,西区,東区,南区	静岡県	
浜北市	浜北区	静岡県	中山間
天竜市	天竜区	静岡県	山間部
舞阪町	西区	静岡県浜名郡	
雄踏町	西区	静岡県浜名郡	
細江町	北区	静岡県引佐郡	中山間
引佐町	北区	静岡県引佐郡	中山間
三ヶ日町	北区	静岡県引佐郡	中山間
春野町	天竜区	静岡県周智郡	山間部
佐久間町	天竜区	静岡県磐田郡	山間部
水窪町	天竜区	静岡県磐田郡	山間部
龍山村	天竜区	静岡県磐田郡	山間部

表 2. 区毎の祭りの確認数(平成 28 年 4 月現在)

区	祭りの数(件)
中区	11
東区	19
西区	41
南区	15
北区	63
浜北区	37
天竜区	176

※合計 362 件

また、合併される中で、継承に関する問題も見え始めてきている。小中学校の統廃合によって、祭りの重要な担い手である子育て世代が都市部に引っ越し、特に山間部においては継承の危機感が高まってきている[4]。表 3 は浜松市における、国・県・市から指定を受けている文化財の一覧である。特に、国や県から指定された文化財は、中山間部・山間部に多くある。さらに、多くの祭りは村人のためであるが、高齢化のため参加することが難しくなり参加者の減少も問題となる。誰も見に来ない状況になれば、主催者にとってもやる意味を見出すことが困難になり、やめる理由となり

得る。神事については、村人のためというより、神に捧げるものであるので必ずしもやめる理由とはならない。

表 3. 無形民俗の文化財指定(浜松市)

文化財名称	地区	指定の種類
寺野のひよんどり	北区引佐町	国指定
川名のひよんどり	北区引佐町	国指定
懐山のおくない	天竜区懐山	国指定
西浦の田楽	天竜区水窪町	国指定
呉松の大念仏	西区呉松町	県指定
滝沢の放歌踊	北区滝沢町	県指定
横尾歌舞伎	北区引佐町	県指定
川合花の舞	天竜区佐久間町	県指定
西浦の念仏踊	天竜区水窪町	県指定
遠州大念仏	中区鹿谷町	市指定
妙功庵観音堂の百万遍念仏と念仏講	北区細江町	市指定
犬居つなん曳	天竜区春野町	市指定
勝坂神楽	天竜区春野町	市指定
滝沢のシノウチ行事	北区滝沢町	国選択
今田花の舞	天竜区佐久間町	県選択
東久留女木の万歳楽	北区引佐町	市認定
雄踏歌舞伎「万人講」	西区雄踏町	市認定
浦川歌舞伎	天竜区佐久間町	市認定
有玉神社の流鏝馬神事	東区有玉南町	市認定
息神社の田遊祭	西区雄踏町	市認定

## 3. 浜松おまつり暦の取り組み

本プロジェクトは、平成 26 年 10 月から平成 27 年 3 月末までに実施されたプロジェクトで、浜松市の祭りに関するデータを収集し、Web カレンダー形式で、Web データベースを構築して、公開した[5,6]。静岡大学情報学部の当研究室 3 年生 6 名と浜松市市民部文化財課が連携して、実施した。

このプロジェクトでは、伝統的な祭りだけにこだわらず、広く行事として行われているイベントについても収集し(ただし音楽関連を除く)、最終的に 525 件収集した(文化財課から提供を受けた 286 件と学生が収集した 239 件)。ほとんどのデータは資料調査をもとに収集した。

これらのデータを元に、データベースの設計を行い、Web カレンダーを実装し、公開している[2]。Web カレンダーのキャプチャ画面を図 1 に掲載する。基本的にカレンダー形式で月毎にその月に開催される祭りが表示される仕様である。また、祭りの矩形を選択すると、祭りの詳細ページを見ることができる。詳細ページでは、名称・住所・日時・地図・概要・写真など基本情報に加え、主催者・開催状況なども閲覧することができる。

## データベースの設計

本システムの全体の ER 図を報告書最後の図 12 に掲載する。データベースの設計において、祭りの開催日を記述する日付のデータは重要となってくる。祭りの開催については、古くは旧暦の月によって決められており、今でも西浦の田楽や舞阪大太鼓祭りなどでは、旧暦を守って開催している。近年ではなかなか職場を休むことができず、ある月の第何土曜日や、最終土曜日、下旬の日曜日といった決め方をしている祭りもある。また、川名のひよんどりなど現在の暦で 1 月 4 日と固定している祭りもある。このように開催日の決め方は様々である。

本プロジェクトでは、日付の設計において、以下のよう

- ・固定日付: 開催日が固定な祭り
- ・変動日付: 第何〇曜日、最後の〇曜日、最初の〇曜日、といった変動だが曜日で確定できる祭り
- ・その他日付: 上記以外の祭り、旧暦開催の祭り、下旬の〇曜日といった祭り、不定期の祭り

Web カレンダーとしては、その年の開催日を表示できることが望ましいが、実質的に毎年更新の必要もあり、全てにおいて確認するのは困難なため、昨年の更新を持って日付の更新は休止している。システム的には該当する年の日付がなければ、汎用的な表示になるため問題はない。また、内部的に日付が計算可能な祭りもあるが基本的に行わない。



図1. 浜松おまつり暦の画面キャプチャ  
(上: カレンダー表示、下: 詳細表示)

## 4. 浜松おまつりアーカイブの取り組み

本事業は、平成 28 年度・平成 29 年度に「みんなのはまつ創造プロジェクト」採択され、浜松お祭りアーカイブの作成を実施した。全点調査を実施したのは平成 28 年度の事業で、具体的な実施方法は以下のとおりである。

- (1) 浜松市の伝統的な祭りの現状調査に関する情報収集(300 件程度)
  - (2) 祭りの詳細調査(10 件程度)
  - (3) 収集した情報の電子化およびアーカイブ化
  - (4) アーカイブを公開するための Web システムの設計・実装・公開・運営
  - (5) 調査した祭りを広報する広報誌の発行と配布
- 以下、調査とデータベース化について述べる。

### 4.1 調査概要

今回の調査では、平成 26 年度の調査で収集された、525 件のうち、産業祭りや月毎で重複しているものを除いて、地域の伝統行事や伝統芸能をともなう祭りに絞り込み、計 362 件の祭りや行事を対象とした。これらの祭りについて基本情報を改めて収集すると同時に、約 20 年経った現在も開催されているか、運営面の現状や実情について調査を行った。

### 調査対象

浜松市内に存在する祭りや行事は 362 件、その内訳は、中区 11 件、東区 19 件、西区 41 件、南区 15 件、北区 63 件、浜北区 37 件、天竜区 176 件である。

### 調査方法

図 2 のように、市内各協働センターを通じ、各地域の祭りや郷土史に詳しい人物や自治会長の連絡先を紹介いただき、郵送調査を行った。また、必要に応じて、電話やメールにて追加で聞き取りを行った。

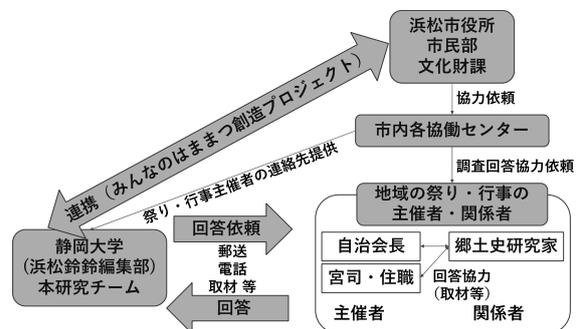


図2. 調査ネットワーク図

### 調査期間

平成 28 年 8 月 1 日～平成 29 年 3 月 31 日

## 調査の設問

設問は、データベース構築用の祭りや行事の基本情報に関する8つ(表4)と、継承状況や後継者不足の有無、今後の開催にあたっての課題等について聞いた7つの質問(表5)で構成した。

表4. 祭りの基本情報の設問

設問No.	祭りや行事の基本情報について
1	祭り・行事の名称
2	開催日
3	開催場所
4	次回平成29年度の開催予定日
5	主催者の名称(保存会等)
6	祭り・行事の代表者のお名前・所属・連絡先(住所、電話番号、メールアドレス等連絡の取りやすいもの)
7	祭り・行事の概要
8	祭り・行事の特徴や伝統

表5. 継承状況等の設問

設問No.	祭りや行事の伝承や運営について
9	現在の祭り・行事の伝承状況を教えてください。当てはまるものに○をつけて下さい。 1.盛ん 2.順調 3.危機
10	祭り・行事の運営に関わっているおおよその人数を教えてください。 ( )人程度
11	後継者不足を感じますか、また、あればその理由も教えてください。 1.継続が危機的な状況 2.深刻な状況 3.少し心配な状況 4.後継者が十分にいる 5.その他( ) 具体的な理由( )
12	今後も祭りや行事を継続するにあたっての課題はありますか。ある場合は、その具体的な内容も教えてください。 1.ある 2.ない 具体的な内容( )
13	今後も祭りや行事を続けていくための活動を行っていますか。行っている場合は具体的な内容も教えてください。 1.行っている 2.行っていない 具体的な内容( )
14	地域の子どもを対象に祭り・行事の指導等の活動を行っていますか。 1.はい 2.いいえ 3.その他( ) あれば具体的な内容も教えてください( )
15	祭り・行事の際に使用する、手順が書かれた手引書や書籍・冊子はありますか。 1.はい 2.いいえ 3.その他( ) あればどんなものがあるか教えてください( )

## 回収状況

平成29年3月31日の時点で、回収数は297件であった。そのうち27件は、新たに見つかったもので、調査対象362件に入っていなかった祭りや行事である。回答率は郵送調査としては82%と高かった。その要因としては、協働センターという公的な機関からの依頼により、関係者に安心感を与えたためと思われる。

回収できなかった92件については、「郵送配布後に返信がなかった」「該当の祭りを知っている関係者が見つけられなかった」の2つのケースがあった。後者については、20年前の調査以降に祭り・行事が廃れてしまったか、名称や内容に変更があり、現在の祭り・行事と関連付け出来なかった可能性が考えられる。

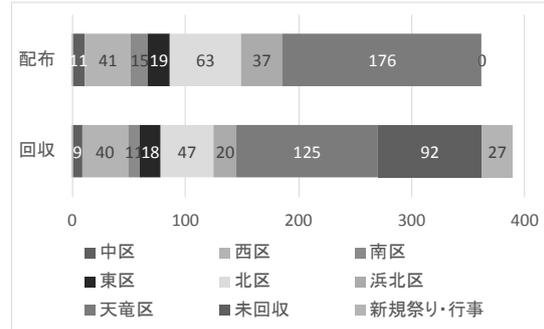


図3. 区ごとのアンケート調査の回答状況

## 4.2 データベース化

これらの調査をもとにして、Web上のデータベースを構築した[3]。図4に「浜松お祭りアーカイブ」と称したWebデータベースの画面キャプチャを掲載する。開催地を表す地図上のプロット表示とカテゴリ検索を基本とする。また、詳細表示、複数の祭りを並べて表示するお祭り比較の機能を持つ。

基本的なデータベース設計については図12と同様である。これに地図上で表示するためのGPS座標と、調査で得られた継承状況に関するデータを持っている。また、Web上では公開していないが、主催者の連絡先に関する情報、調査状況、最寄りの協働センター、アンケート時に自由記述で記載された内容についてもデータは持っている。



図4. 浜松お祭りアーカイブの画面キャプチャ (上:地図上プロット表示、下:カテゴリ検索)

## 5. 継承状況調査の結果

ここでは、継承状況に関する表5の7つの質問について、その回答結果を報告する。

### (1) 現在の祭り・行事の伝承状況について

市内全体で見ると、回答の約半数の祭りや行事において、開催されていた。しかし、前回市が行った調査以降、約20年の間に、30件の祭りや行事が開催されなくなっていた。うち29件は山間地域であった。一方、伝承状況が盛んと回答のあった祭りの多くは「大太鼓」「花火」「屋台」といった見どころがあり、観光も意識した規模の大きな祭りであることが分かった。

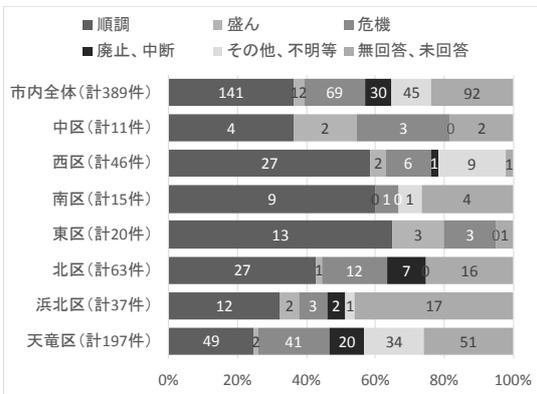


図5. 伝承状況について

### (2) 祭り・行事の運営に関わるおおよその人数

回答の割合は、10人～50人程度での運営が多いことが分かった。天竜区は他の区に比べ、10人程度以下の少人数で行われている祭りや行事が多い。これは、天竜区における運営者の高齢化と若い世代の人口減少が原因と思われる。しかし、他の情報も絡めて見ると、運営人数と伝承状況の関連は、必ずしも強くないことが分かった。規模を小さくすることで、運営人数が少なくても安定して開催を続けられるような工夫をしているところがあった。

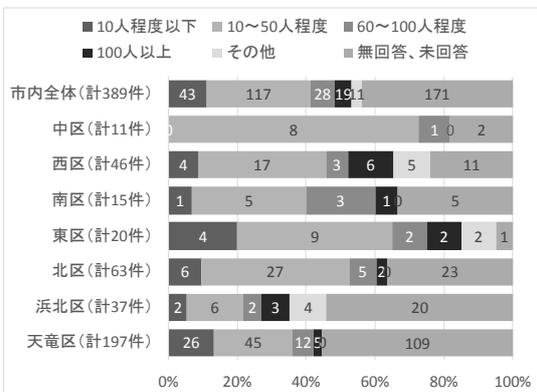


図6. 運営に関わるおおよその人数

### (3) 後継者不足の有無

全体の約7割は、多少なりとも後継者不足を感じていた。その理由として、「若者の人口減少」「運営者の高齢化により開催が困難になった」「宗教離れ」「祭りや地域の行事に対する住民の意識変化」等が多く挙げられた。

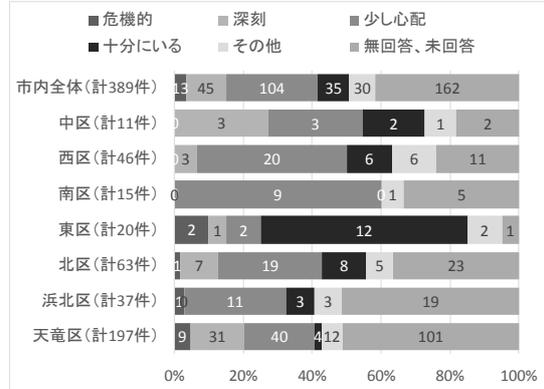


図7. 後継者不足の有無

### (4) 今後も祭りや行事を継続するうえでの課題の有無

全体で3割以上の祭りや行事について、課題はあるとの回答があった。具体的には、「若者世代の加入と後継者の発掘、育成」が多く挙げられた。他にも、開催場所へのアクセス面に対する課題も見られた。山間地域においては、車道がない山道を徒歩で行かなければならない等、高齢者の参加が減少する要因があった。

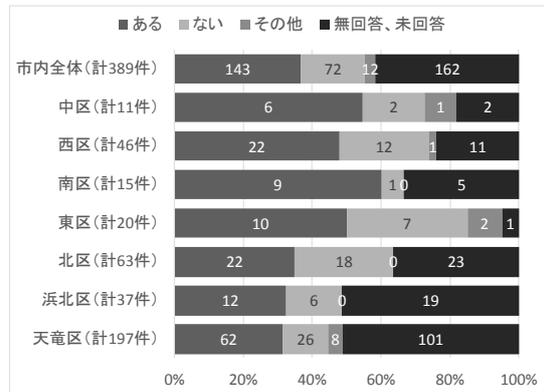


図8. 課題の有無

### (5) 今後も祭りや行事を続けていくための活動の有無

全体の約3割が、祭や行事を続けていくために何らかの活動を行っていた。例えば、内部、外部に向けた宣伝活動に力を入れる、若者世代が減っている山間地域では、地域から出ていった若者に声をかけ参加を呼びかける等があった。町外の人に協力を要請したり、運営の年齢層を上げたりすることで継続を図っている組織もあった。

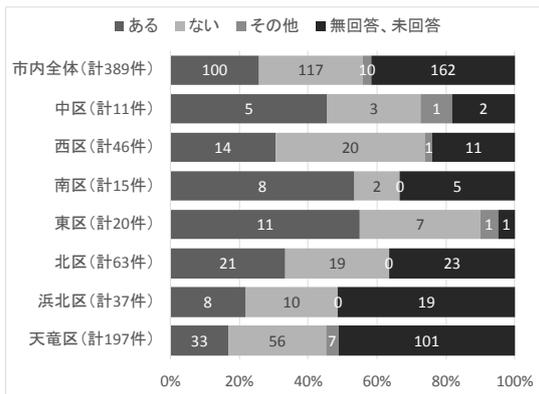


図 9. 継続のため活動の有無

### (6) 地域の子供を対象にした指導等活動の有無

全体的に指導等の活動を行っているところは多くなかった。特に少なかった浜北区、天竜区については、そもそも地域に子供がないという回答が多かった。指導等の具体的な内容は、お囃子や舞の練習を行うのがほとんどであった。中には祭事で使用する道具(松明等)の準備を手伝ってもらおうというもあった。

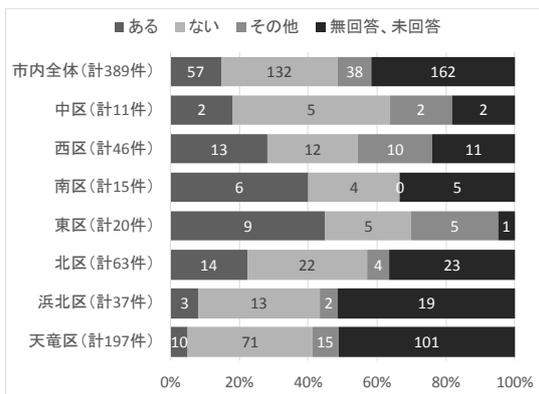


図 10. 子どもを対象にした活動の有無

### (7) 祭り・行事に使用する手引書や資料等の有無

手引書等の冊子や資料を作成し残しているところはとても少なく、天竜区、浜北区においては0という結果であった。また、あると回答のあったところについても、進行手順のみ、舞や歌についての作法についてのみで、祭り全体について記録してあるものを持っているところは、ほぼなかった。多くが、年度が変わる際に主に口頭で引き継ぎをするとのことであった。

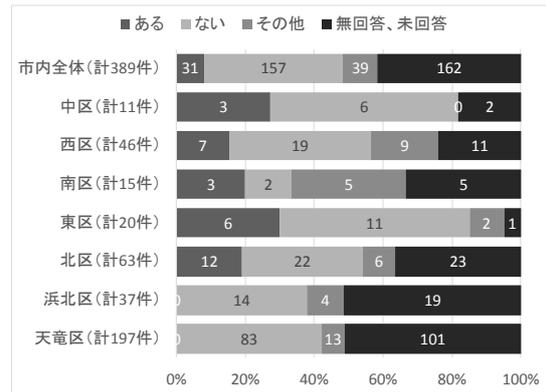


図 11. 手引き書や資料などの有無

### (8) 調査を終えて

平成 29 年 3 月 31 日の時点で、浜松市内では少なくとも 222 件の祭りや地域の行事が開催されている。しかしそのほぼ全てにおいて、高齢化、若者を主とする人口減少がうかがえる。また、祭りや地域の行事は、各地域の神社で行われることが多いが、今回の調査で神社関係者と地域住民のつながりの希薄化を感じた。近年の宗教離れの加速に伴い、祭りや神事に対する地域住民の意識も、祖先・神を祀り、祈願・感謝するものから、単なる年間行事の一つに変わってきた。それによって祭りは、神社関係者による神事と地域住民による余興とで別れ、その繋がりが弱まってきているのである。聞き取りの中で、この繋がりの希薄化を問題視し、氏子が自治会に協力を求め、結束の強化、問題の共有を行うことで、祭りの廃止の危機を脱したという意見があった。

手引書等資料のない祭りや行事が多いという調査結果は、今回のデータベース化の意義を感じた。祭りが廃れないことが一番ではあるが、たとえ廃れてしまったとしても記録に残しておくことで、後日復活出来る可能性がある。

## 6. まとめ

本報告では、浜松市における「祭り・神事・行事」についてのデータベースの構築と継承状況の調査について報告した。全点調査を目標に実施したが、まだ 3 割は現状を把握できておらず、特に、浜松市の中心部である中区については、人間関係の希薄化で、市役所からのトップダウン的な調査では、(すでに途絶えている可能性もあるが)市民レベルで残されている行事などは確認できていない。市民レベルのネットワークや足を使った調査を必要としている。また、神社などで独自に(もしくは秘事として)行っている祭りや神事もあり、その点についてもトップダウン的な調査では漏れがあることが分かっている。これらについては静岡県神社庁などの協力も必要となってくる。

今回の調査で、少子化のためすでに廃止になっていたり、行事を略したり、開催時間を変えたりといった、急激な変化が起こっていることが分かった。この点についても、情報として残していくアーカイブ化が強く必要とされている。

## 謝辞

調査および取材にご協力賜りました関係者の皆様に厚く御礼申し上げます。本報告に関連するプロジェクトを推進してくれた、研究室の学生諸君らに感謝する。本研究の一部は科研費基盤研究(C)15K01147の助成を受けたものである。本研究は浜松市の平成28年度・平成29年度の「みんなのはままつ創造プロジェクト」の補助を受けたものである。

## 参考文献

- [1] 浜松市例規集,「浜松市民俗芸能の継承及び振興に関する条例」,[http://www1.g-reiki.net/hamamatsu/reiki\\_honbun/o700RG00001696.html](http://www1.g-reiki.net/hamamatsu/reiki_honbun/o700RG00001696.html), (参照 2018-02-09).
- [2] 静岡大学情報学部杉山岳弘研究室,「浜松おまつり暦」, <http://www.hama365.info/matsuri/>, (参照 2018-02-09).
- [3] 静岡大学情報学部杉山岳弘研究室,「浜松お祭りアーカイブ」, <http://www.hama365.info/archive/>, (参照 2018-02-09).
- [4] 村井梨乃,「浜松市北部の伝統的祭りにおける継承に関する取り組みの調査」,2016年度卒業論文,静岡大学情報学部情報社会学科 (2017)
- [5] 2015年4月18日 中日新聞 18版「浜松の祭り 有名無名 525件 静大生ら写真、動画で紹介」
- [6] 2015年4月18日 静岡新聞 県内総合(22)「浜松の祭り ウェブで発信 カレンダー形式 525件 静大生が取材、開設」

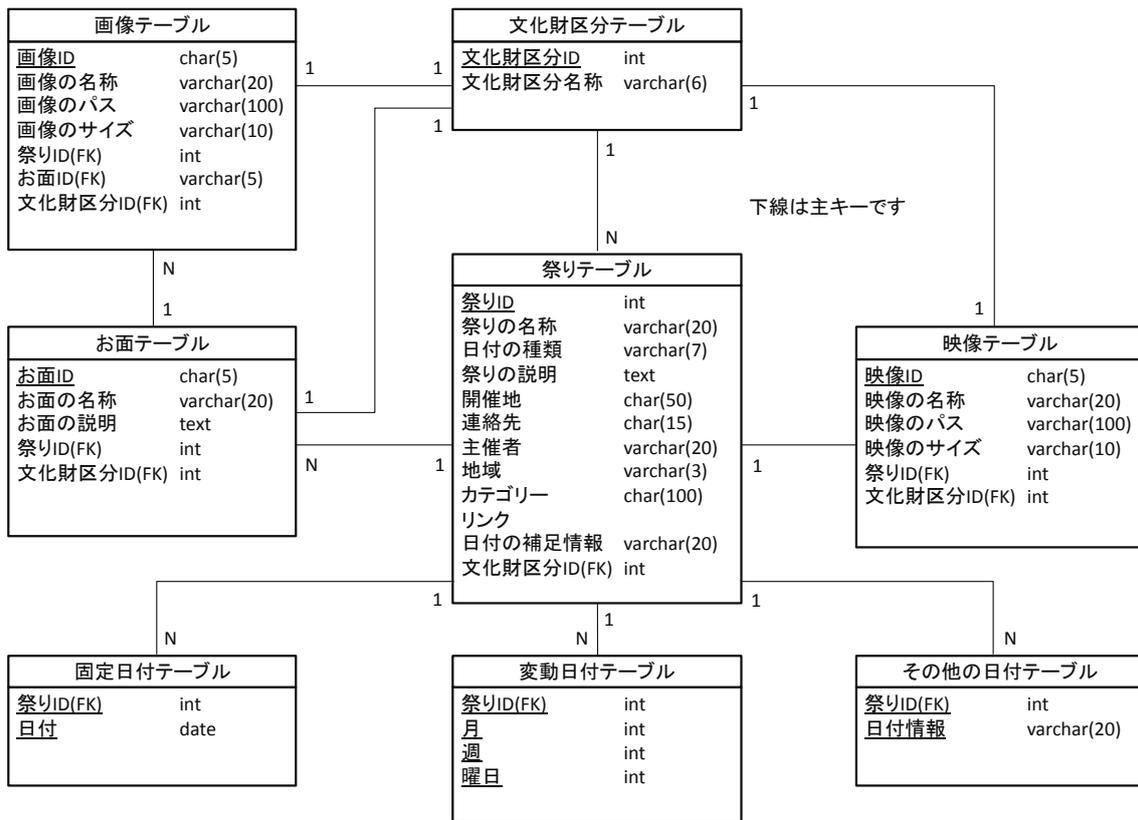


図 12. 浜松おまつり暦 (Web カレンダー) のデータベースの ER 図

# 語の出現傾向による『源氏物語』第一部各巻の分類

## A Quantitative Study of "The Tale of Genji" via a Statistical Analysis of the Word Frequency

土山 玄

Gen Tsuchiyama

一橋大学 森有礼高等教育国際流動化センター, 東京都国立市中 2-1 第3研究館  
Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo

**概要:** 『源氏物語』は平安時代に成立した 54 巻から構成される長編物語である。一般に『源氏物語』は三部構成であると考えられており、第一部は第 1 巻「桐壺」から第 33 巻「藤裏葉」までが該当し、これら 33 巻は「紫上系」と「玉鬘系」という 2 つの群に分類されるという見解がある。この見解は「紫上系」の登場人物は「玉鬘系」の諸巻にも現れるが、「玉鬘系」に初出の人物は例外なく「紫上系」の諸巻には現れないという事実に基づいている。しかし、第一部 33 巻が 2 群に分類されるとする見解を支持する根拠は登場人物の出現状況の他にない。そこで、本研究では語の出現傾向を統計的に分析することで、計量的な観点より両群の文体的特徴に相違が認められるか検討を加えた。分析の結果、出現頻度の高い語の出現頻度、出現頻度の高い機能語の出現頻度が両群に間において相違することが明らかになった。

**Abstract:** "The Tale of Genji" is one of the most famous classical works of Japanese literature and one of the oldest full-length stories. Most studies on Japanese literature have considered that the story comprises 54 volumes that are categorized into 3 parts, with volumes 1-33 categorized under the first part. However, there is an idea that the order of these 33 volumes is different from the order of the number of volumes and there is a possibility that these volumes are classified into 2 groups: "Murasaki no Ue Goup" and "Tamakazura Group." Furthermore, there are some objective datasets that support this idea. Herein, we statistically analyzed the writing style of the aforementioned groups by performing multivariable analysis and found a statistical difference between the writing style of these groups.

**キーワード:** 源氏物語, 計量文献学, 多変量解析, 主成分分析

**Keywords:** The Tale of Genji, stylometry, multivariable analysis, principal component analysis

### 1. はじめに

『源氏物語』は平安時代に成立した長編物語である。この物語は 54 巻から構成され、一般に『源氏物語』は三部に分割されると考えられている[1]。第一部には第 1 巻「桐壺」から第 33 巻「藤裏葉」までが属し、第二部は第 34 巻「若菜上」から第 41 巻「幻」の 8 巻が、第三部は第 42 巻「匂宮」から第 54 巻「夢浮橋」の 13 巻が属すとされる。この第一部の 33 巻には成立順序が現行の巻序と異なるという見解があり、これら 33 巻は「紫上系」と「玉鬘系」という 2 つの群に分類されると論じら

れている[2]。「紫上系」及び「玉鬘系」の分類は表 1 に示す通りである。このように第一部 33 巻が 2 群に分類されるという見解の根拠として、登場人物の出現状況があげられる。紫上系に登場した人物は玉鬘系においても出現するが、反対に玉鬘系において初出となる登場人物は例外なく紫上系に登場することはない。武田 (1954) によれば、このような事実に基づき、紫上系の 17 巻が成立した後に玉鬘系が成立し年立に従い紫上系に挿入されたと論じられている[2]。しかし、第

一部 33 巻が 2 群に分類されるとする見解を支持する根拠は登場人物の出現状況の他にない。

そこで、本研究では『源氏物語』第一部の成立過程を解明するために、語の出現頻度について統計的に分析を行う。文学的文章のテキストデータを用いた計量的な研究は計量文献学と称される。文章の計量分析では、主に著者の文体に関わる習慣的特徴を統計的に解析する。文体という概念は多様であるが、計量文献学においては文体とは計数可能な記述形式のことであり、その内容は文字や語の頻度、語や文の長さなどの文章を構成する量的な要素である。

このような文章の計量分析において、著者について議論の余地がある文章を対象に、計量的な手法による研究は従来から行われている。計量的に著者の推定あるいは識別を行う場合、文中において語彙的意味ではなく文法的機能を担う助詞や助動詞などの機能語が広く分析に用いられ、欧米諸語で記述された文献や日本の近現代の文献を対象に、その成果が報告されている。一方で、テキストの電子化が容易ではないことから、日本の古典文学作品を対象とした研究は十分に展開されているとは言えない。これに加えて、『源氏物語』などの日本の古典文学作品の多くはオリジナル原稿が散逸しており、書写によってのみ受け継がれており、このため著者の特徴的な文体が希釈されている可能性が予想される。しかし、土山・村上 (2011) では、『うつほ物語』と『源氏物語』を対象とし、現代文や欧米文学と同様に古典文においても、語の出現率、あるいは語の頻度について計量分析を行うことで著者の識別が可能であることを論じている[3]。

このような著者不詳の文献を対象とした著者の識別や推定を目的とした計量的な研究では、機能語と称される文中において語彙的意味ではなく文法的機能を担う語彙が分析項目として取り上げられることが一般的である。日本語の場合は、助詞や助動詞などが機能語に該当し、このような機能語が計量分析に取り上げられる背景には、名詞や動詞などの語彙的意味を担う実質語に比べ、機能語は研究対象となる文献の

表 1 紫上系の諸巻及び玉鬘系の諸巻

紫上系		玉鬘系	
01桐壺	13明石	02帚木	24胡蝶
05若紫	14滯標	03空蝉	25蛩
07紅葉賀	17絵合	04夕顔	26常夏
08花宴	18松風	06末摘花	27篝火
09葵	19薄雲	15蓬生	28野分
10賢木	20朝顔	16関屋	29行幸
11花散里	21少女	22玉鬘	30藤袴
12須磨	32梅枝	23初音	31真木柱
	33藤裏葉		

内容、すなわちストーリーの影響を受けにくく、出現する語彙に記述内容に起因すると考えられる出現傾向の偏りが生じにくく、著者の識別に適していると考えられることがある。

よって、本研究ではこのような計量文献学の方法を用いて、『源氏物語』の第一部について、紫上系と玉鬘系との間に計量的な相違が認められるのか検討を加える。

## 2. 関連研究

第一部の成立過程については、すでに計量的な検討が加えられている。村上・今西 (1999) は『源氏物語』の全文を電子化したテキストデータを用い、多変量解析を行った初の本格的な研究である。村上・今西 (1999) では、各巻の助動詞の出現率を求め数量化 III 類を行っている。分析の結果、紫上系は玉鬘系よりもむしろ第二部の 8 巻と助動詞の出現傾向は類似していることを明らかにした。従って、第 1 巻「桐壺」を起筆の巻と仮定するのであれば、紫上系が成立した後に第二部が成立し、その後に玉鬘系が成立した可能性を論じている[4]。

次いで、小野 (2015) は村上・今西 (1999) と同一のデータを用いて、分散安定化変換を行い、階層的クラスタ分析を用いて『源氏物語』の成立過程について検討を加えている。分析の結果、『源氏物語』の諸

巻は紫上系及び第二部、玉鬘系及び宇治十帖という2群に分類されることを報告した[5]。

また、土山 (2016) では助動詞以外の品詞についても巻別の出現率を求め主成分分析を行っている。その結果、紫上系と玉鬘系との間において出現傾向が異なっている品詞は助動詞のみであることを指摘し、紫上系の諸巻は玉鬘系の諸巻に比べて使役・尊敬の助動詞が頻出していることを明らかにした。またこれに加えて、動詞の未然形に接続する受身・尊敬・可能・自発の助動詞である「る」は紫上系に、「らる」は「玉鬘系」に多く出現することを報告した。

### 3. 分析

#### 3.1 データ

本研究では青表紙本系の大島本を主な底本とする『源氏物語語彙用例総索引 自立語編』[6]及び『源氏物語語彙用例総索引 付属語編』[7]を電子化したテキストデータを分析に利用した。『源氏物語語彙用例総索引』は源氏物語の本文すべてについて、形態素解析を行ったものである。すべての語に、表記形、表記形仮名読み、終止形、終止形仮名読み、品詞コード、活用形コード、意味コードなどの情報が付与されている。なお、単語認定については、『源氏物語大成索引篇』[8]の単語認定基準に準拠している。

#### 3.2 方法

本研究では語の n-gram を分析に採り上げた。n-gram は文中において隣接する n 個の要素を 1 つの単位とする。よって、単語の n-gram は隣接する n 個の単語を 1 つの単位として頻度を集計した特徴量である。n = 1 のときは unigram、n = 2

のとき bigram、n = 3 のときは trigram と称され、本研究では unigram 及び bigram を分析に用いた。『源氏物語』の冒頭の文章である「いつれの御時にか女御更衣あまたさふらひ給けるなかにいとやむことなききはにはあらぬかすくれて時めき給ありけり」を例とすると、unigram は「いつれ (代名詞)」「の (助詞)」「御時 (名詞)」「に (助詞)」「か (助詞)」などの各語の頻度を集計する。他方、bigram では「いつれ (代名詞)–の (助詞)」「の (助詞)–御時 (名詞)」「御時 (名詞)–に (助詞)」「に (助詞)–か (助詞)」という語の隣接共起した組の頻度が集計される。

また、文章の計量分析では品詞別に語の unigram を集計し、品詞毎に分析を行うことが多い。しかし本研究では、まず品詞別に語を集計せずに出現頻度上位の unigram 及び bigram を用いて分析を行った。次いで、先にふれたように、研究対象となる文章の内容の影響を受けにくいと考えられる機能語の unigram 及び bigram を用いて分析を行った。機能語とは文章中に

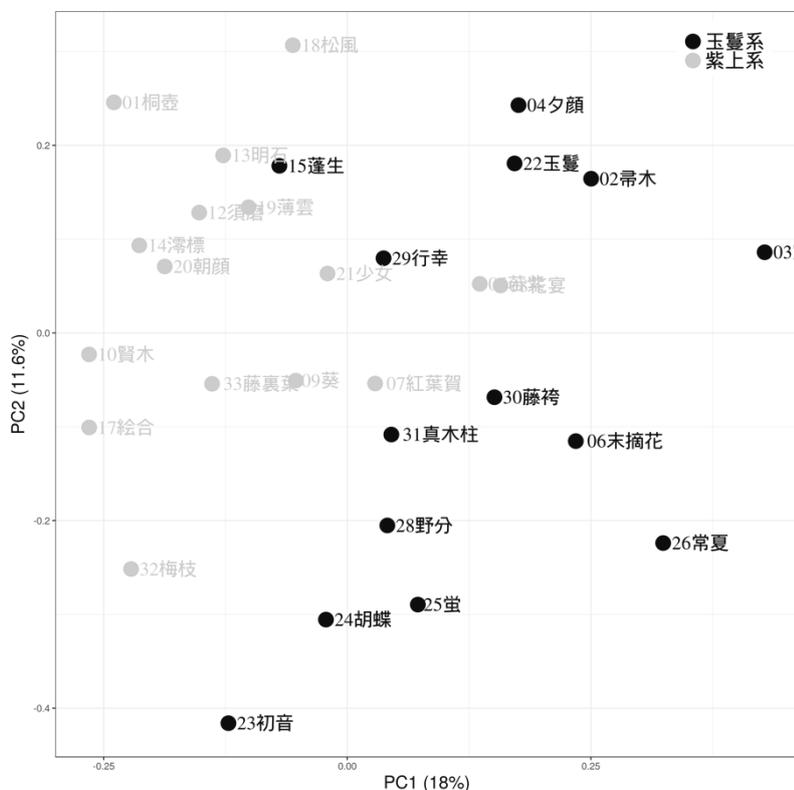


図1 unigram の出現頻度上位 50 変数についての主成分分析の結果

において語彙的意味を担わず文法的機能を担う語であることから、本研究では補助動詞、助詞、助動詞の3品詞を機能語として、頻度を集計した。上掲の『源氏物語』冒頭の文章を例とすると機能語は「の(助詞)」「に(助詞)」「か(助詞)」「給(補助動詞)」「ける(助動詞)」が該当する。また、bigram は隣接する語であることから機能語の bigram は「に(助詞)-か(助詞)」「給(補助動詞)-ける(助動詞)」などとなる。このような品詞及び単語の n-gram の集計を巻別に行った。ただし、『源氏物語』の各巻の延べ語数は一様ではないことから、各 n-gram

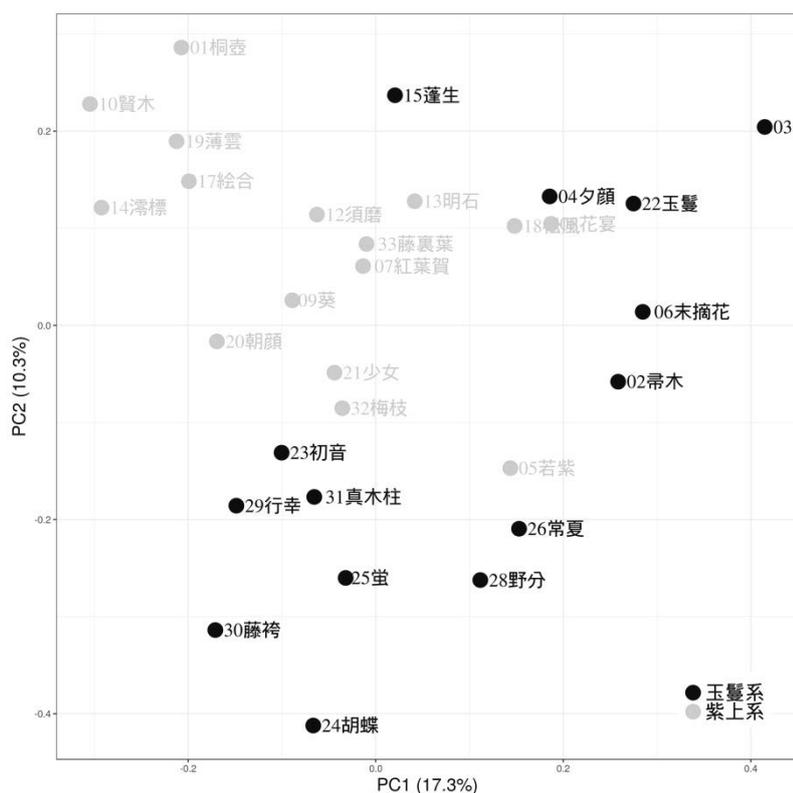


図2 bigram の出現頻度上位 50 変数についての主成分分析の結果

の頻度を分析には用いず、n-gram の総度数に対する割合、すなわち各 n-gram の出現率を求め、これを分析に用いた。また、n-gram の集計に際しては、補助動詞と助動詞の活用する 2 品詞は終止形に直した。このように集計した n-gram の出現率について主成分分析を行った。なお、主成分分析では相関係数行列を用いた。

また、本研究においては延べ語数が 1000 語を下回る第 11 巻「花散里」(724 語)、第 16 巻「関屋」(934 語)、第 27 巻「篝火」(653 語) を分析から除外した。

### 3.3 分析結果

まず、unigram の出現頻度上位 50 変数について主成分分析を行った。上位 50 変数は頻度が 422 以上の語が該当し、上位 50 変数までの累積度数は総度数の 54.0%である。この出現頻度上位 50 変数の unigram には名詞や動詞と言った出現傾向が物語の内容に影響を受ける語彙も含まれている。しかし、上位 50 変数に含まれる名詞は「こと」「ひと」「ほど」「ころ」「さま」も

の「よ」の 7 語であり、一般的かつ抽象的な語であると考えられることから、物語の筋によって出現頻度が大きく影響を与えられないことが推測される。動詞についても、「あり」「おもふ」「おぼす」「みる」「す」などの語が含まれ、副詞は「いと」が含まれるのみである。図 1 は主成分分析によって求められた第 1 主成分と第 2 主成分の主成分得点の散布図である。図 1 において、紫上系に属する第 5 巻「若紫」及び第 8 巻「花宴」の 2 巻が玉鬘系に接近し、玉鬘系に属する第 15 巻「蓬生」及び第 29 巻「行幸」の 2 巻が紫上系に接近して位置しているが、紫上系の諸巻と玉鬘系の諸巻は第 1 主成分において概ね分離され付置されていることが分かる。紫上系においては「若紫」と「花宴」に第 7 巻「紅葉賀」を加えた 3 巻の第 1 主成分得点が正になるが他の巻は負になり、その一方で、玉鬘系の諸巻は「蓬生」、第 23 巻「初音」、第 24 巻「胡蝶」の 3 巻の第 1 主成分得点は負になるが他の巻は正になる。なお、第 1 主成分

の寄与率は 18.0%、第 2 主成分の寄与率は 11.6%である。

次いで、bigram の出現頻度上位 50 変数について主成分分析を行った。上位 50 変数は頻度が 173 以上の bigram が該当し、上位 50 変数までの累積度数は総度数の 9.4%である。図 2 は主成分分析の結果であり、図 1 と同様に紫上系に属する第 5 巻「若紫」及び第 8 巻「花宴」が玉鬘系の諸巻に接近して付置されている。また、unigram の出現頻度上位 50 変数についての分析結果とは異なり、図 2 においては第 18 巻「松風」も玉鬘系の諸巻に接近して位置している。しかし、紫上系の諸巻は玉鬘系の諸巻とは異なり主に図中の第 2 象限に付置されていることから、bigram の分析においても第一部は 2 群に分類され得ると考えられる。なお、第 1 主成分の寄与率は 18.0%、第 2 主成分の寄与率は 11.6%である。また、上述したように bigram の出現頻度上位 50 変数までの累積度数の総度数に対する割合は 9.4%と低いが、変数を増加させても分析結果は大きく変わらない。

次に、機能語について分析を行った。先にふれたように、本研究では補助動詞、助詞、助動詞を分析に用いた。機能語の unigram の出現頻度上位 50 変数について主成分分析を行った。上位 50 変数は頻度が 179 以上の語が該当し、上位 50 変数までの累積度数は総度数の 98.3%である。主成分分析の結果は図 3 に示す通りであり、紫上系の属する多くの巻は第

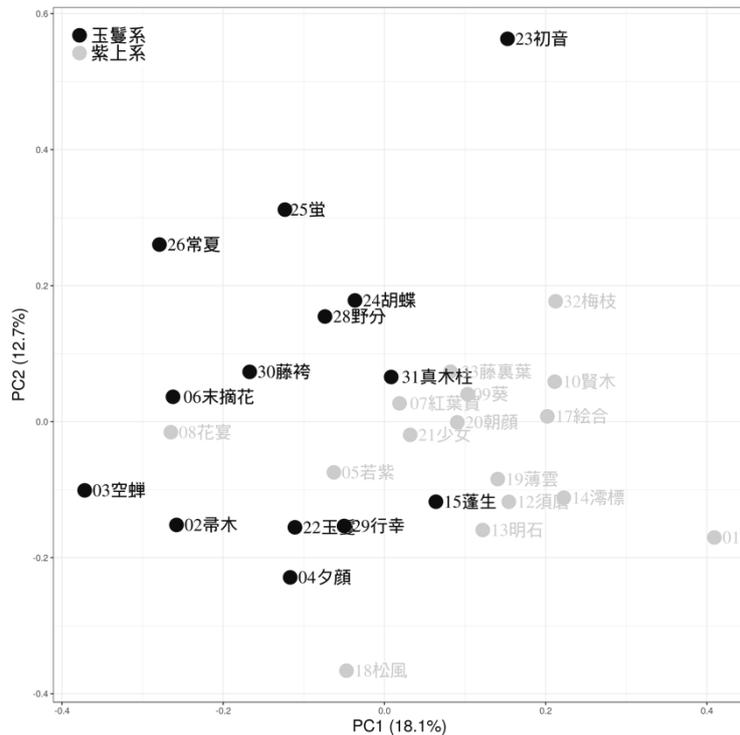


図 3 機能語 unigram の出現頻度上位 50 変数についての主成分分析の結果

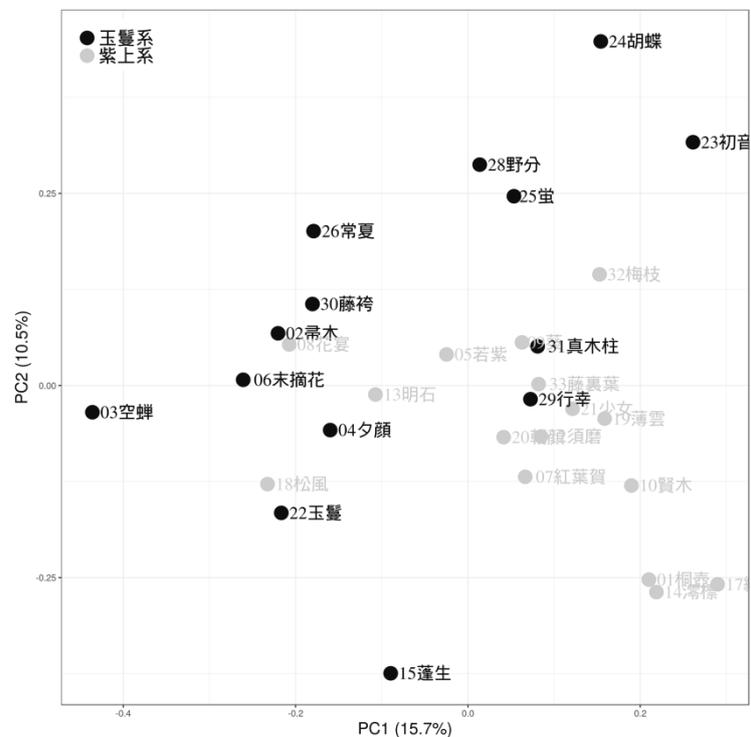


図 4 機能語 bigram の出現頻度上位 50 変数についての主成分分析の結果

1 主成分の正の領域に、玉鬘系の諸巻は負の領域に付置されており、第一部の諸巻は概ね紫上系と玉鬘系で分離して位置していると考えられる。なお、第 1 主成分の寄与率は 18.1%、第 2 主成分の寄与率は 12.7%である。

最後に、機能語の bigram の出現頻度上位 50 変数について主成分分析を行った。上位 50 変数は頻度が 110 以上の bigram が該当し、上位 50 変数までの累積度数は総度数の 49.9%である。図 4 は主成分分析の結果であり、紫上系の多くの巻は主に図中の第 4 象限に位置している。よって、機能語の bigram の分析結果から紫上系の諸巻と玉鬘系の諸巻は異なる量的傾向を有していると考えられる。ただし、紫上系の第 8 巻「花宴」、第 13 巻「明石」、第 18 巻「松風」は玉鬘系の諸巻と混在し、玉鬘系の第 29 巻「行幸」及び第 31 巻「真木柱」は紫上系の諸巻と混在している。なお、第 1 主成分の寄与率は 15.7%、第 2 主成分の寄与率は 10.5%である。

#### 4. 考察

作家が文章を執筆する上でどのような語を多く用いるか、ということはその作家の習慣的特徴の現れであると考えられる。本研究において用いた出現頻度上位の語の unigram 及び語の bigram は作家が文章を執筆する際の習慣的特徴であり、文体的形式的特徴を規定する要素の 1 つであると思われる。本研究において検討を加えた項目は『源氏物語』にあらわれる出現頻度上位の unigram 及び bigram、そして機能語に限定した unigram 及び bigram である。

各 n-gram の出現頻度上位 50 変数について主成分分析を行った結果、どの分析結果においても紫上系の諸巻と玉鬘系の諸巻は一部の巻が混在して付置されるものの概ね異なる傾向を有していると考えられる。従って、計量的な判断に基づくならば、文体的形式的特徴は相違していると考えられる。

#### 参考文献

[1] 池田亀鑑. 源氏物語の構成 (新講源氏物語 (上)

所収). 至文堂, 1951.

- [2] 武田宗俊. 源氏物語の研究. 岩波書店, 1954.
- [3] 土山玄・村上征勝. 源氏物語と宇津保物語における語の使用傾向について. じんもんこん 2011 論文集, 2011(8), 125-132, 2011.
- [4] 村上征勝・今西祐一郎. 源氏物語の助動詞の計量分析. 情報処理学会論文誌, 40(3), 774-782, 1999.
- [5] 小野洋平. 源氏物語成立論の統計科学的再考察: 村上・今西(1999)を中心に. 計量国語学, 29(8), 296-312, 2015.
- [6] 上田英代・村上征勝・今西祐一郎・樺島忠夫・上田裕一. 源氏物語語彙用例総索引—自立語編一. 勉誠出版, 1994.
- [7] 上田英代・村上征勝・今西祐一郎・樺島忠夫・上田裕一. 源氏物語語彙用例総索引—付属語編一. 勉誠出版, 1996.
- [8] 池田亀鑑. 源氏物語大成 索引篇, 中央公論社, 1985.

# 中尾佐助スライドデータベースと大学リポジトリとの連携に向けて Toward System Cooperation between Sasuke Nakao Slide Database and Institutional Repository

小島 篤博, 青木 茂樹, 泉 正夫, 宮本 貴朗

Atsuhiko Kojima, Shigeki Aoki, Masao Izumi, Takao Miyamoto

大阪府立大学大学院 人間社会システム科学研究科, 大阪府堺市中区学園町 1-1

Osaka Prefecture University, 1-1, Gakuen-cho, Nakaku, Sakai, Osaka

**概要:** 大阪府立大学では、故中尾佐助博士がフィールド調査において撮影した資料をデータベース化した中尾佐助スライドデータベースを開発、公開している。2000年の公開以来、数度のシステム更新を経て現在に至るが、基本的な設計は開発当初から大きく変更されておらず、長期的な運用が懸念されつつある。一方、近年ではデジタル資料の機関リポジトリへの集約により、永続的なアーカイブ化が進められている。本報告では、中尾佐助スライドデータベースのこれまでの運用状況とともに、大学リポジトリとの連携を考慮した今後の展望について紹介する。

**Abstract:** Library and science information center of Osaka Prefecture University maintains a slide image database taken by the late Dr. Sasuke Nakao. Since 2000 when the system was made to the public, it has been updated several times as the basic structure remains unchanged. In recent years, such research materials are being gathered on persistent digital archives in the institutional repositories. This paper describes the present state of the Sasuke Nakao slide database as well as a prospect of the system cooperation between the database and the University's institutional repository.

**キーワード:** 中尾佐助, 照葉樹林文化論, 画像データベース, 機関リポジトリ

**Keywords:** Sasuke Nakao, theory of evergreen forest culture, image database, institutional repository

## 1. はじめに

大阪府立大学学術情報センターには、照葉樹林文化論[1]の提唱者として名高い故・中尾佐助大阪府立大学名誉教授の旧蔵書、雑誌、スライド、遠征アルバム、フィールドノート、スクラップブックなど膨大な研究資料(中尾佐助コレクション)が保存され、大阪府立大学の特色あるコレクションの一つとなっている。これらのうち写真やスライド、フィールドノートなどについては、スキャナにより画像データとして読み取り、Web上で閲覧可能な中尾佐助スライドデータベースとして公開されている。

現在公開の対象となっているのは、中尾資料のうち22,000枚のスライドであり、これに撮影場所や日時などのメタデータを付与した上で、キーワードや場所、日時などにより検索することができるようになっている。学術的・文化的な研究資料をデジタルアーカイブ化し、

インターネット上に公開する試みは他にもなされているが、本システムの特徴は、研究者が個々の画像ごとに独自のキーワードやコメントを付与する機能を有することである。大阪府立大学では1997年よりこのデータベースの開発を進め、公開から約20年を経た現在までに幾度かシステム更新を行ってきたが、基本的な設計は当初のものをそのまま踏襲している。研究機関が独自に開発したシステムは、研究者個人の努力によるものが多く、異動や退職により運用の継続が難しくなるなどの懸念を抱えており、人文系データベースの分野においてもメンテナンスを中断せざるを得なかった例も聞かれる。

一方、近年では、大学などの研究機関が所蔵する論文や図書、研究資料等を機関リポジトリとして集約する動きが進んでいる。中尾佐助資料に関しても、資料の一部であるフィールドノートが大阪府立大学学術情報



図1 中尾佐助スライドデータベース

リポジトリ OPERA に収容されている。機関リポジトリに収容された資料は、永続的な URI を与えられ将来に渡って利用可能であることが期待できる。したがって、前述の独自開発型データベースと機関リポジトリを横断的に連携することで、長期的にもデジタルアーカイブとして利用できることを担保しつつ、その時代に応じた独自機能の追加を実現することができると考えられる。

本論文では、中尾佐助スライドデータベースのシステム構成について紹介するとともに、大学リポジトリとの連携により、デジタルアーカイブとしての継続性と独自機能の実現の両立を目指した将来構想について述べる。

以下、2では中尾佐助資料の概要について紹介し、続いて3では現行システムの概要について説明する。4では、データベースの構成とシステムの具体的な機能について説明し、5で大学リポジトリの連携について展望する。最後に6でまとめる。

## 2. 中尾佐助資料について

大阪府立大学が保管する中尾佐助コレクションは、旧蔵書 3,500 冊、雑誌 60 タイトル、著書 18 点、学术论文 88 点、新聞・雑誌等の記事 552 点、スライド約 28,000 点、遠征アルバム 77 冊及び写真ネガ、スクラップブック 94 冊、記録ノート 116 冊、フィールドノート 19 冊等からなり、中尾佐助コーナーにおいて閲覧に供している[2]。このうち、スライドフィルムは 1955 年から 1984 年の海外調査において撮影されたものである。遠征アルバムには、1935 年頃から 1950 年代前半のカラースライドにするまでものが白黒写真で残されている。

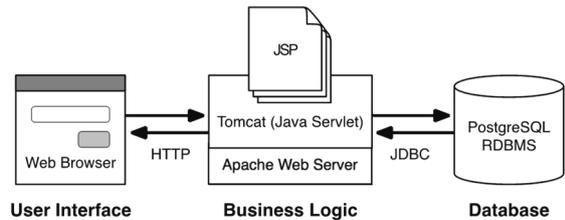


図2 システム構成図

写真帳の写真の番号とネガの番号が併記しており、個々のスライドを特定できるようになっている。

大阪府立大学学術情報センターでは、これら中尾資料のデジタルアーカイブ化を進めている。まず、1999 年より第 1 期開発として、中尾が 1958 年にブータンのフィールド調査を行った際のスライド画像 1,300 枚を対象とし、個々のスライドにキーワードや場所、日時などを関連付けることで、キーワードのほか地理的、時間的條件から目的の画像を検索することができる画像データベースを構築し、公開してきた[3][4]。これに続く第 2 期開発として、2006 年度より、残りのカラースライド画像についてもキーワード付与などの作業を行い、第 1 期公開分と合わせて約 22,000 枚の画像データベースとして 2009 年より公開している(図 1)。後述する研究者支援機能は、この第 2 期において実装したものである[5]。

さて、これらのスライドフィルムの形態としては、個々のスライドのマウントに、1 枚ごとに識別番号が付与されている。この識別番号をもとに、フィルムごとに分類整理され、保管されている。例えば、“N58-54-12”は、1958 年に撮影された、54 番目のフィルムの 12 コマ目となる。この他、スライドの情報を得る手がかりとしては、中尾自身の著作である『秘境ブータン』[6]、『ヒマラヤの花』[7]、『ブータンの花』[8]があり、フィールドノート(3 冊)とブータンのアルバム(12 冊)が揃っていた。また、フィールドノートには、キャンプ地と標高、キャンプ地からキャンプ地への移動行程や天候等が明記されており、スライドを特定する際に有用な情報を得ることができた。

## 3. 現行システムの構成

### 3.1 システムの設計

現行の中尾佐助スライドデータベースは、以下の機能を有している。

- (1) ユーザ登録・管理
- (2) さまざまな条件によるスライドの検索
- (3) コメントやキーワードの付与
- (4) ブックマークの付与

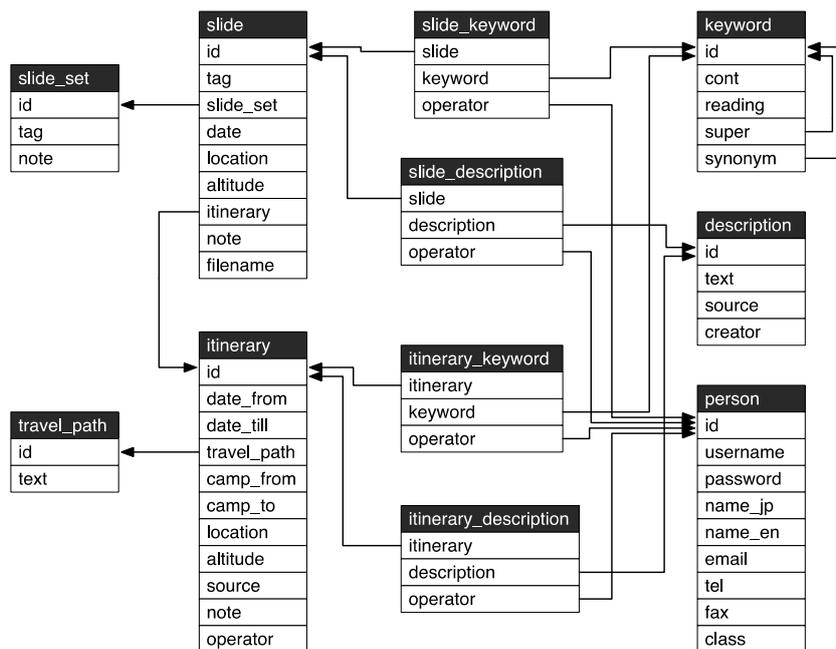


図3 データベースのE-R図

表1 スキーマ一覧

スキーマ	内容
slide	スライドの基本情報
slide_set	スライドセット単位の情報
itinerary	フィールド調査の旅程
keyword	キーワード情報
description	文献のテキスト
slide_keyword	スライド=キーワード対応表
slide_description	スライド=テキスト対応表
person	登録ユーザ情報
comment	ユーザが付与したコメント
bookmark	ユーザが付与したブックマーク

まず、ユーザごとに作業履歴等を保存するため、研究者向け機能の利用にあたってはユーザ登録をいただいている。ユーザは、利用開始時にパスワードによる認証を行い、以後の作業はそのユーザに関連付けて記録される。なお、研究者専用機能を利用しない場合は、ユーザ認証なしで検索や画像表示などの基本的な閲覧機能が提供されている。

スライドの検索は、キーワード、日付、番号、および一覧からの選択のいずれかの方法により行うことができる。スライドを特定した後、必要があれば、そのスライドにコメントやブックマークを付与することができる。コメントは、スライドの内容について自由記述形式で記入ことができ、公開の可否も選択可能である。また、ブックマークに登録しておけば、あとで簡単にアクセスすることができる。

以上のような要件を考慮してシステム構成を検討した結果、図2に示すような3階層型システム(3-tier system)として実現することとした。これにより、一般ユーザからのスライド検索や、特定ユーザによるデータ編集作業などはWebインタフェースにより行い、そのためのビジネスロジックはWebサーバ上に実装し、さらに基本データ構造やそれらの関連性はRDBMSに格納することができる。

Webサーバの構成はApacheとTomcatにより実装し、動的なページ生成にはJSP(JavaServer Pages)を利用している。JSPでは、サーバ側の処理をJava言語で記述できるため、Webからの要求に応じて動的にSQLを組み立てて、RDBMSにクエリを投入するなど、きめ細かな処理を行うことが可能である。バックエンドのRDBMSとしてはPostgreSQLを採用しており、JDBC(Java Database Connectivity)を介してTomcat上のJavaと接続している。また、ODBC(Open Database Connectivity)を介してMicrosoft Accessなどのデータベース・フロントエンドと接続することで、容易にデータの編集が可能となっている。

### 3.2 データベースの定義

スライド画像をキーワードから検索するには、適切なキーワードをスライドに対応付ける必要がある。また、キーワードやコメントを付与した作業者を記録するためには、ユーザ情報についてもデータベースに格納する必要がある。本研究では、これらの項目を組み合わせ、

### キーワード検索

検索するキーワードを入力してください:  
(前のボタンを押すとキーワードの一覧が表示されます。)

AND ↓

AND ↓

AND ↓

AND ↓

AND ↓

キーワード検索: 部分一致 ↓

アルファベット: 大文字・小文字区別なし ↓

クリア 検索

キーワードには、漢字を含む日本語、全角カタカナによる読み、および半角アルファ

図 4 キーワード検索画面



日付: 1958-07-12  
場所: チョモラリ・ゴンバ  
キーワード: ダイオウ, 高山植物, 葉, 中尾佐助, 植物, 巨大, 黄白色

### 新規コメントを追加

コメントを公開する  
※公開するを選択した場合、コメントとともに個人の氏名も公開されます。

ダイオウの花と中尾佐助

クリア 送信

図 6 コメント編集画面

### スライド表示

スライド番号: N62-10-07



カテゴリ: 東ネパール 1962  
日付: 1962-06-15  
場所: グンサ  
キーワード: 白, 雪, 山

### 編集メニュー

このスライドに関する情報の編集が行えます

▶ ブックマークを付与	▶ キーワードを付与	▶ コメントを付与
▶ 日付・場所を編集	▶ 別のスライドを選択	▶ メニューへ戻る

図 5 スライド編集画面

1 つの検索システムとして機能するようデータベースのスキーマ設計を行った。本システムのスキーマの関連を表した E-R 図を図 3 に、また、スキーマの一覧を表 1 に示す。

ここで、slide は個々のスライドの基本情報(撮影日時、場所、画像ファイル名など)、keyword はキーワード情報(キーワード、英表記、同義語へのリンクなど)、また person は登録ユーザ情報(ユーザ名、パスワード、氏名、連絡先など)を保持している。

### 3.3 スライドの検索

キーワードによるスライドの検索は、ユーザが入力したキーワードと適合するレコードを keyword テーブルから検索し、そこから slide\_keyword を辿ることで目的の slide を選択する。また、複数のキーワードを指定する場合は、それぞれ選択されたスライドの集合に対し、積集合や和集合をとることができる。

実際のキーワード検索画面を図 4 に示す。キーワードとして利用できる文字は、漢字、ひらがな、全角カタカナ、英数字である。また、個々のキーワードに対して、前方一致、後方一致、部分一致、完全一致、アルファベットの大小文字の区別も指定できるようになっている。キーワードを指定して「検索」をクリックすると、該当するスライドのサムネイル画像が一覧表示され、さらにサムネイル画像を選んでクリックすると、スライド画像とともにキーワード、コメントなどの付随情報が表示される(図 5)。

### 3.4 データの編集

スライドの基本情報として、キーワードやコメントを追加することができる。キーワードを付与するためには、まず対象のスライドを選択する。例えば、先述の検索により選択されたスライド画面の下には編集メニューが表示され、ブックマーク、キーワード、コメントを付与できるようになっている。ここで「キーワードを付与」というリンクをクリックすると、キーワード編集画面が表示され、新たなキーワードの追加や既存のキーワードの削除ができる。なお、既存のキーワードは、付与者自身または管理者によってしか削除できないように制限されている。

新たに入力したキーワードが、既存の keyword のテーブルに存在しなければ、新たに keyword のレコードを追加する。同時に、slide\_keyword のテーブルにレコードを追加し、それぞれ slide と keyword との関連付けを行う。

コメントの付与についても、キーワード同様、対象スライドを選択したのち編集メニューから行うことができる。「コメントを付与」をクリックすると、図 6 のようなコメント編集画面が表示される。この画面は、コメントを公開す



索引					
頁数	摘要	頁数	摘要	頁数	摘要
6	0900-1000	26	11,990	25	13
7	1800-1900	70	18,235	19	47
6	1300-1500	70	11,705	25	6
7	2200-2400	70	18,235	19	8
6	1000-1200	70	11,705	25	6
		70	18,235	19	8
6	1500-1600	70	11,705	25	6
7	0000-0100	70	18,235	19	8

図7 大阪府立大学リポジトリ

るかどうかのチェックボックスと、自由記述形式でコメントを入力できるテキストボックスから構成されている。それぞれ入力し、「送信」ボタンをクリックすれば、対象スライドにコメントを付与することができる。

## 4. 大学リポジトリとの連携

### 4.1 現行システムの課題

ここまで現状のシステム構成について述べてきたが、将来的な課題として、いかに本システムのサービスを長期的に継続していくかという懸念がある。研究資料のデジタルアーカイブは永続的なアクセスが期待される一方、独自開発されたシステムでは OS やサーバソフトウェアのサポート期限や、ミドルウェアの開発停止など寿命がある。多大な労力を費やして構築したデータベースも、一定期間ごとにシステムを更新し続けなければ、やがて利用できなくなってしまう。

本システムも、当初は Linux 2.2 上に Java ベースの Web サーバである Jigsaw 2 を基盤に Java Servlet として実現し、DB サーバは Windows NT 4.0 上の Oracle 8 に接続する 2 サーバ構成で構築した。しかしながら、Jigsaw の開発が停止し Java ベースの Web アプリケーション基盤は Apache Tomcat プロジェクトに移行したため、第 2 期では RedHat Enterprise Linux 5、Apache 2.0、Tomcat 6、そして RDBMS は PostgreSQL 8 に更新した。現行システムでは CentOS 7、Apache 2.4、Tomcat 7、PostgreSQL 9 に更新している。Web アプリケーションのトレンドは現在も変化しており、PHP のほか Ruby

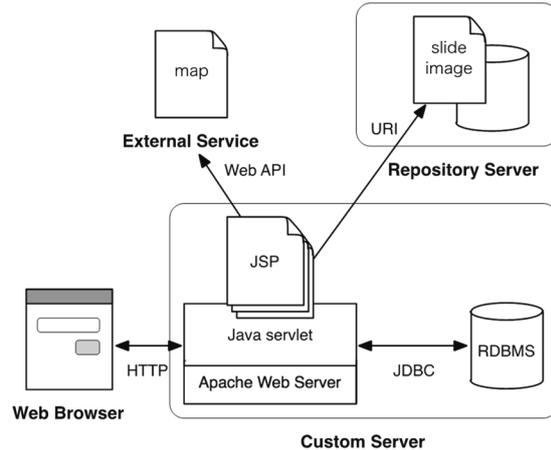


図8 リポジトリとの連携

on Rails や Python が普及し、更には開発当初には存在しなかったマッシュアップ型の様々なサービスが登場してきている。

蓄積したデータを将来に亘って安定的に提供し続けるためには、より汎用的・公共的な基盤を利用するのが確実であると考えられる。近年、大学や国立研究機関で集約が進められている機関リポジトリは、まさに研究資料を蓄積する目的で開発されたものである。大阪府立大学においても、オープンソースのデジタルアーカイブシステムである DSpace[9]を採用した学術情報リポジトリ OPERA を運営している[10]。このリポジトリには中尾コレクションの一部も収容されており、図 7 にフィールドノート(1958年8月16日-10月26日):ブータンの掲載ページを示す。DSpace は多くのデータ形式をサポートしており、画像データをそのメタデータとともに蓄積・管理できるため、本システムのスライド画像を移行することができれば、長期的に安定したデータの管理とサービスを維持することが期待できる。一方、検索方式は画一的であり、基本的にはテキストのパターンマッチングのみとなってしまうため、現在本システムが提供しているような地図から撮影場所を絞り込んだり、撮影日時で検索したりするなどの独自の機能は実現できない。

### 4.2 将来の構想

そこで、本システムの将来構想として、画像データやそれに付随する基本情報は大学リポジトリに収容し、独自の検索機能は別システムとして分離して、2 サーバを連携することでシステムを実現することを検討している。図 8 にそのシステム構成を示す。現行のシステムに蓄積されているスライド画像データとその基本情報は、リポジトリに収容可能なデータ形式に変換して移行する。その際、個々のスライドに付与されたキーワ

ードは、テキスト形式で出力してメタデータの一部とする。リポジトリがサポートしない独自の検索、例えば地図や日時からのスライド検索はカスタム検索サーバ上でを行い、検索結果のスライド画像のみリポジトリ上の URI にリンクする形でページを生成する。

これにより、スライド画像を含む基本的なデータは大学リポジトリ上で長期的な参照を担保しつつ、独自機能は別サーバに分離し、万が一サービスを継続できなくなった場合でも基本的なデータだけはリポジトリ上に残すことができる。

更には、カスタム検索サーバを実現する際、地図上の位置の参照は Google Maps API など外部サービスを利用したり、キーワードは Linked Data[11]や RDF に準拠することで汎用性を高めたりするなど、独自開発が必要な部分を徐々に縮小していくことも考えられる。

以上の構想に基づいて、現在はシステムの基本設計を行なっている段階であり、基本機能の分離を実現したパイロットサーバを開発する予定である。

## 5. おわりに

大阪府立大学が保管する中尾佐助資料を Web 上で検索可能な中尾佐助スライドデータベースと、その継続的な利用を目指した大学リポジトリとの連携について述べた。本システムは以下の URL で公開しており、研究者登録をしていただければ、キーワードやコメントを付与することもできる。関連分野の研究者にご利用いただければ幸いです。

<http://nakao-db.center.osakafu-u.ac.jp/>

## 謝辞

本研究において、データベースの要件および機能改善について有益なご助言いただいた山口裕文大阪府立大学名誉教授、ならびにデータベースの運営に日々ご協力いただいている大阪府立大学学術情報室の皆様には感謝の意を表します。

## 参考文献

- [1] 中尾佐助. 中尾佐助著作集 第 IV 卷 照葉樹林文化論. 札幌, 北海道大学出版会, 2006.
- [2] 大阪府立大学総合情報センター編. 中尾佐助文献・資料総目: 照葉樹林文化論の源流. 堺, 大阪府立大学総合情報センター, 1997.
- [3] 富坂敏子, 小池利栄子, 小島篤博, 石井敬三, 宮本貴朗, 山野美贇子. 中尾佐助資料スライドデータベースの構築と利用者間情報共有への展望. レコード・マネジメント, No.41, pp.8-18, Nov. 2000.
- [4] 小島篤博, 宮本貴朗, 富坂敏子, 小池利栄子, 石井敬三, 山野美贇子. 貴重文化資料デジタルアーカイブの Web による検索・共有システム

の構築. 電子情報通信学会技術報告, ET2001-31, pp.61-68, Jul. 2001.

- [5] 小島篤博, 青木茂樹, 宮本貴朗. 中尾佐助資料公開のための画像データベースの構築と運用. 公開シンポジウム「人文科学とデータベース」, Vol.16, pp.1-8, 2010.
- [6] 中尾佐助. 秘境ブータン (岩波現代文庫). 岩波書店, 2011.
- [7] 中尾佐助. ヒマラヤの花. 東京, 毎日新聞社, 1964.
- [8] 中尾佐助, 西岡京治. ブータンの花. 東京, 朝日新聞社, 1984.
- [9] “DSpace Official Site”. <http://www.dspace.org/>
- [10] “大阪府立大学学術情報リポジトリ OPERA”. <http://repository.osakafu-u.ac.jp/dspace/>
- [11] T.ヒース, C.バイツァー (武田英明監訳). Linded Data: Web をグローバルなデータ空間にする仕組み. 近代科学社, 2013.

## 郷土食による地域理解支援システム「もちマップ」の オープンデータ化についての考察

### A Consideration on Open Data Conversion for Implementing Local Foods-base Region Understanding Support System “Mochi-Map”

河村郁江<sup>1</sup>, 伊藤宗太<sup>2</sup>, 伊藤孝行<sup>1</sup>, 白松俊<sup>1</sup>

Ikue Kawamura, Sota Ito, Takayuki Ito, Shun Shiramatsu

1 名古屋工業大学大学院 情報工学専攻, 名古屋市昭和区御器所町

Nagoya Institute of Technology University, Gokiso-cho, Showa-ku, Nagoya, Aichi

2 インフォ・ラウンジ合同会社, 神奈川県横浜市都筑区茅ヶ崎中央

Info Lounge LLC, Chigasaki chuo, Tsuzuki-ku, Yokohama, Kanagawa

**概要:** 郷土食は、古くから地域の伝統的な産物を活用している食べ物である。各地の郷土食をよく知る事は、地域の理解や知見につながる。当研究では、2016年から郷土食による地域理解支援システムを試作している。本研究では、2017年10月に行われた共通語彙基盤ハッカソンをきっかけに、システムで利用しているデータをLOD(Linked Open Data)にしたことを述べ、今後のシステムやデータの活用について考察する。

**Abstract:** Since local food have been made of local products, view point of local food culture is effective for local understanding. We have been prototyping a regional understanding support system by local food from 2016. We participated in a hackathon on utilizing IMI (Infrastructure for Multilayer Interoperability) at October 2017. Since this event, we converted our data about local food into linked open data (LOD) based on the IMI vocabularies. This paper discusses our perspective about this system and utilizing data in the future.

**キーワード:** 郷土食, WEB マップ, 情報の構造化, Linked Open Data

**Keywords:** Local food, Web map, Structuring information, Linked Open Data

#### 1. はじめに

郷土食は地域の産物を活用した食べ物である。郷土食をよく知ることは、地域の風土や文化、自身の住む地域や他地域との違い、および時代の変化について知ることであり、地域の理解や知見につながる。

しかし、昔から各地域で作られ続けてきた郷土食は、近年の便利な食生活の中では手間がかかることや、地域の少子化のため作り伝えられなくなり、自分の住む地域の郷土食について深く知る人は減って来ている。また、郷土食は地域ごとの解釈が異なることや、時代による食生活の変化等により、元々の成り立ちは分かりにくくなっている。そのため2016年の秋から制作中のWEBマップでは、日本地図からもちの分布を見る事で、郷土食への興味や理解を深めることを目指した[1][2]。図1は本研究が目指すユースケースの概要図である。

本研究で作成しているシステムは、郷土食情報としてもちの情報を扱うため、名称を「もちマップ」とした。

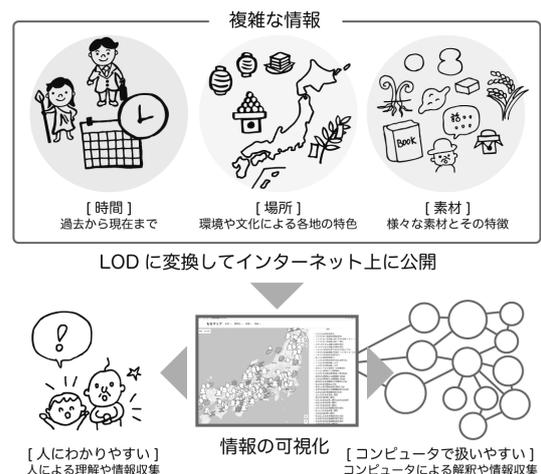


図1 本研究が目指すユースケースの概要

もちを主題として扱う理由を下記に述べる。

- ・ 地域の生活文化と深く結びつき、ハレの食から、日常の食、および保存食として食べられて来た。

- ・ 時代に伴い食生活が変化する中でも、長く食べ続けられている。
- ・ 日本全体で食べられ、使用素材が地域の特色と結びついている場合が多い。

このような理由から、もちは身近であり、郷土食の一例として分かりやすく、地域理解の一助となると考えた。この研究では対象を、郷土食や地域の文化に興味がある人や、学びたい人とする。我々は、このようなもちの情報を整理し、視覚化するため、「もちマップ」を開発してきた。

我々は、2017年10月8,9日に開催された“実践！共通語彙基盤ワークショップ・ハッカソン”[3]において、本システムで利用しているもちデータを Linked Open Data (LOD) と呼ばれる形式のオープンデータにした。LOD とは、様々な組織が公開するオープンデータを関連付け、組み合わせることで活用できるようにする技術的枠組みである。また、LOD 化するにあたって、共通語彙基盤を用いた。共通語彙基盤とは、行政などが保有する公的なデータを対象とし、経済産業省所轄の情報処理推進機構 (IPA) が標準化作業を試みている語彙セットである。これに基づいて LOD 化したことにより、データの相互運用性 (Interoperability) が向上し、他のオープンデータと組み合わせた分析が可能になると期待される。

データをオープンデータ化するために何をしたのか、また今後どのようにオープンデータを利用して行くかについて、これまでの制作の問題点も含めて考察する。

## 2. これまでの開発と課題

我々は2016年の冬からもちマップを開発してきた。最初に制作した従来のもちマップをバージョン1とし、2017年のものをバージョン2とする。本研究はバージョン3となる。以下にこれまでどのような開発をし、どのような課題があるかを述べる。

### 【もちマップ バージョン1】

もちマップバージョン1では、もちデータを書籍[4][5]やインターネット上の情報、および実際に人から聞いた話などを元に作成した。システムの主な機能として、以下の4つの機能を実装し、システムの位置や素材、および調理法などの属性を可視化した。

- (1) 検索窓
- (2) 属性選択ボタン
- (3) 背景とレイヤーをコントロールするチェックボックスパネル
- (4) 時間軸スライダー

日本地図全体上に表示された情報や属性を見ることが、郷土食の特徴や地域の特色がわかれば、地域理

解が深まると考えた。図2は従来のもちマップの主な機能である。



図2 従来のもちマップの主な機能

もちマップの評価を行った所、郷土食理解が深まったという意見が得られた。一方で、バージョン1での課題として下記の内容が指摘された。

- (1) もちの全体像だけでなく、各もちの詳細情報ページを作成する
- (2) 複数人によるもちデータの入力
- (3) より利用しやすく、分かりやすいUIへの改善
- (4) 時間軸機能の改善
- (5) 地域区分や統計情報の工夫

### 【もちマップバージョン2】

その後、2017年にバージョン1で出た課題の中で、下記を試みた。

- ・ 複数人によるもちデータの入力
- ・ 各もちの詳細情報ページの作成

バージョン1ではローカルで読み込んだもちデータを、データベース化した [6]。図3はもちマップバージョン2のデータベース構成である。

データベースの正規化を行う際に、初めてもちデータの語彙の検討をした。図書館でよく使用されている日本十進分類法を見ると、「餅」は「民俗」と「料理」に分類されている。まず「民俗」の語彙で調べた所、国立歴史民俗博物館による民俗語彙データベース[7]が存在したが、参考に留めた。食の語彙の論文[8]によると、「食を学として取り上げる場合、予想される学術領域は図に示すような広範なものとなろう」とあり、図には50種類以上の分野が記載されている。また、Google Scholar[9]で「郷土食」を検索した所、栄養学 約375件、経済学 約373件、社会学 約293件、民俗学(民族学) 約168件、地理学 約159件、調理学 148件、歴史学 約75件、農学 約66件、考古学 約54件、国文学 約5件などがあり、郷土食のみでも広い範囲で扱われているため、ある分野に基準を置くことはせず、どの分野からも利用できる Dublin Core[10]をもち

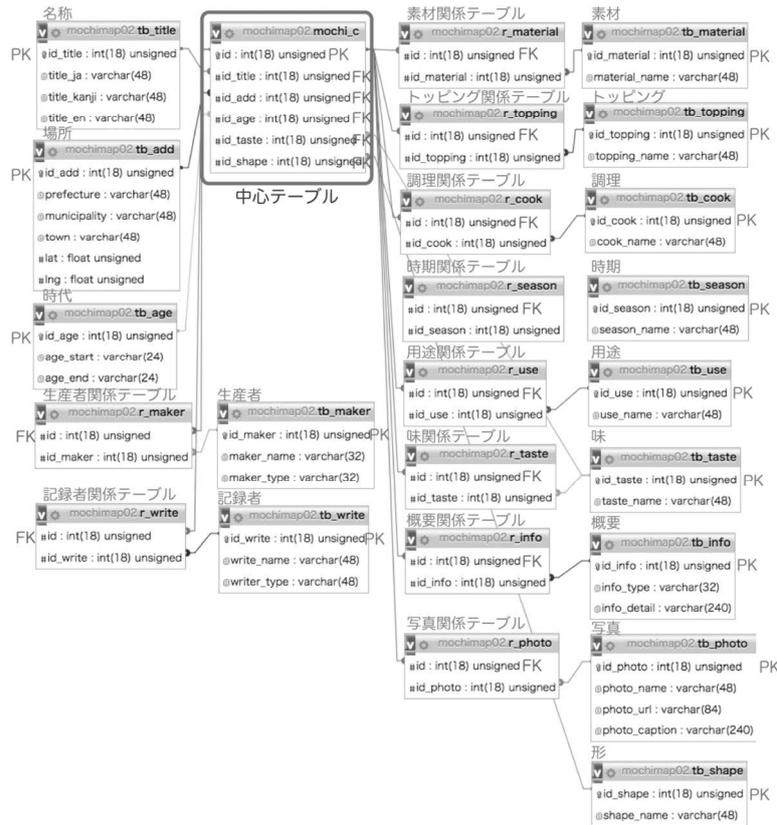


図3 もちマップデータベース構成

```

<section itemscope itemtype="http://schema.org/Recipe">
  <div id="wrap">
    <?php if (isset($error)) : ?>
      <p class="error"><?= h($error); ?></p>
    <?php endif; ?>
    <h1 itemprop="name" class="f2 bk01"><?= h($row['title']); ?> | <?= h($row['title_kanji']); ?></h1>
    <div itemprop="image" itemscope itemtype="http://schema.org/ImageObject" class="p_center">
      " width="80%" class="p_center" />
      <div itemprop="caption" class="p_center"><?= h($row['image_caption']); ?></div>
    <div itemprop="description"><?= h($row['description']); ?></div>
    <dl>
      <dt>素材1</dt><dd itemprop="recipeIngredient"><?= h($row['ingredient01']); ?></dd>
      <dt>素材2</dt><dd itemprop="recipeIngredient"><?= h($row['ingredient02']); ?></dd>
      <dt>素材詳細</dt><dd itemprop="description"><?= h($row['ingredient_detail']); ?></dd>
      <dt>トッピング1</dt><dd itemprop="recipeIngredient"><?= h($row['topping01']); ?></dd>
      <dt>トッピング2</dt><dd itemprop="recipeIngredient"><?= h($row['topping02']); ?></dd>
      <dt>トッピング詳細</dt><dd itemprop="recipeIngredient"><?= h($row['topping_detail']); ?></dd>
      <dt>味</dt><dd itemprop="name"><?= h($row['taste']); ?></dd>
      <dt>調理法</dt><dd itemprop="cookingMethod"><?= h($row['cook']); ?></dd>
      <dt>形</dt><dd itemprop="name"><?= h($row['shape']); ?></dd>
      <dt>形詳細</dt><dd itemprop="description"><?= h($row['shape_detail']); ?></dd>
      <dt>用途</dt><dd itemprop="name"><?= h($row['use01']); ?></dd>
      <dt>用途詳細</dt><dd itemprop="description"><?= h($row['use_detail']); ?></dd>
      <dt>都道府県</dt><dd itemprop="spatialCoverage"><?= h($row['prefecture']); ?></dd>
      <dt>市町村</dt><dd itemprop="spatialCoverage"><?= h($row['municipality']); ?></dd>
      <dt>町域</dt><dd itemprop="spatialCoverage"><?= h($row['town']); ?></dd>
      <dt>時期</dt><dd itemprop="name"><?= h($row['season']); ?></dd>
      <dt>いつからあるか</dt><dd itemprop="name"><?= h($row['start_m']); ?></dd>
      <dt>いつまで存在するか</dt><dd itemprop="name"><?= h($row['end_m']); ?></dd>
      <dt>情報元の種類</dt><dd itemprop="citation"><?= h($row['referring']); ?></dd>
      <dt>情報元</dt><dd itemprop="citation"><?= h($row['referring_detail']); ?></dd>
      <!-- <div><?= h($row['opinion']); ?></div> -->
      <dt>投稿時間</dt><dd itemscope itemtype="http://schema.org/Date" itemprop="commentTime"><?= h($row['time_post']); ?></dd>
      <dt>記入者</dt><dd itemprop="author"><?= h($row['inputter']); ?></dd>
    </dl>
  </div>
</section>

```

図4 もちマップを Schema.org にあてはめた例

データに適用した。Dublin Core は情報資源の組織化のために作られ、語彙が共通認識となるよう設計された基本語彙セットである。しかし、実際に当てはめると、Dublin Core は書籍を想定して作られた規格であるため、郷土食データを当てはめるには無理があった。そのため代わりに Schema.org[11]を使用した。

schema.org はコンピュータにもわかる形式で意味を付ける構造化マークアップのための語彙であり、Google や Microsoft、および Yahoo! が共同で進めている。検索エンジンはこのマークアップを利用し、質が高い検索結果を出すことができるため、ユーザは必要なページに辿り着きやすくなる。schema.org の型の中で食に関するものは、Thing/CreativeWork/Recipe(レシピ)、Thing/Event/FoodEvent(食イベント)および Thing/Organization(または Place)/LocalBusiness/FoodEstablishment(食品関係) などがあるが、郷土食としてのもちを表現するのに合う語彙がなかったため、元々ある語彙を使用した。

Schema.org は独自の拡張も可能である。図 4 はもちマップを Schema.org にあてはめた例である。バージョン 2 においては、このように様々な語彙の検討をした。(2) 複数人によるもちデータの入力については、各もちのページを生成するフォームを作成し、インターネット上に公開した[12]。公開当初にフォーム入力のユーザーテストを行った(2017年12月、20代から40代の男女30名、名古屋近郊のIT系コミュニティイベントにて)。入力ミスを減らすため、基本的には選択式の入力にしたが、手入力の項目では、正しく入力されない例が30件中4件あった。また、「素材」を聞く項目では、「わからない」が30件中2件、「情報源」の項目では「人から聞いた」が30件中10件、「その他」が6件、「WEBの情報」が6件であり、「書籍」という回答は0件であった。以上から、複数人によるもちデータの入力自体は実現できたが、一般向けに入力してもらうデータの信頼性が低いため、フォームのUIを改善することや、管理側での確認など、入力情報の収集・管理方法の検討が必要であることがわかった。

**3. もちデータを Linked Open Data に**  
共通語彙基盤のハッカソン[3]において、もちマップの新しいシステムを作成した(システム制作:伊藤宗太)。これをバージョン3とする。図5はもちマップバージョン3の画面である。

今回の大きな変更点は、1つ目に、もちデータを共通語彙基盤を基にしたデータモデル記述(DMD)にすることで、同時にデータが RDF 形式になり、そのデータを公開 SPARQL エンドポイントに置いたことで LOD (Linked Open Data) になったことである。2つ目には、同じもちデータを利用して、各もちの詳細ページが閲

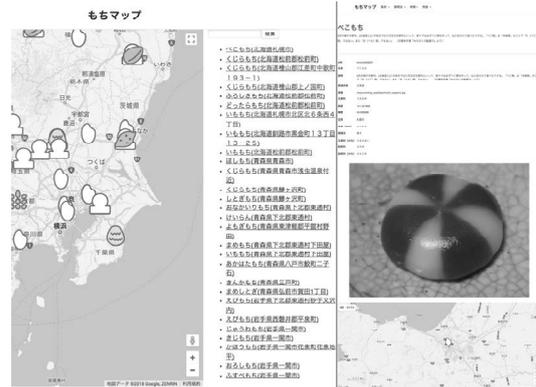


図5 もちマップバージョン3

大カテゴリー	素材	調理方法	時期	用途
• 素材	• すべてのもち	• 焼く	• 正月	• 日常食
• 調理法	• もち米	• 茹でる	• 春	• 間食
• 時期	• もち米+うるち米	• 煮る	• 夏	• お祭り
• 用途	• うるち米	• 炊く	• 秋	• お供え物
• 県別	• 小麦粉	• 練る	• 冬	• 保存食
	• 雑穀(ヒエ、アワなど)	• 干す		• 正月
	• 堅果	• その他		
	• さつまいも			
	• じゃがいも			
	• さといも			
	• 根(クス、ワラビなど)			
	• 草			
	• その他			

図6 従来のもちマップの分類項目

左: トップページ, 右: 各もちのページ例

覧できるようになったことがある。以下にどのような流れで LOD にしたかを説明する。

本章ではもちデータをオープンデータ化した流れを述べる。図6は従来のもちマップの分類項目である。

(1) 共通語彙基盤(IMI)

共通語彙基盤[13]は、共通で使われる語彙と、語彙同士の関係を示す仕組みで構成された、オープンデータの相互運用性を高めるためのフレームワークである。コンピュータ間のデータ通信を円滑かつ確実に行うことが目的であり、一つの語の代表的な意味や構造の概念、および他の概念との関係などを明確にした概念の集合を指して用いる。コンピュータでの処理を難しくさせている用語の問題(類似用語、重複用語、分野で意味が異なる同一語など)を解決するため、概念的な同一性を保証する形で「用語」を定義している。この基準に合わせるために、表形式のデータを DMD(Data Model Description)に基き作成する。DMD はデータの構造や項目、各項目の値の範囲や書式などを説明するものであり、DMD 作成ツールを使用することで、人間にわかりやすく、コンピュータでのデータ交換に適した形式および LOD に適した RDF 形式のデータとして公開することもできる。

もちデータを共通語彙基盤に合わせるに当たり、一段階目に元々独自のシステムに合わせて作成していた不要な項目を整頓し、不必要な項目は削除した。共通語彙基盤のコア語彙バージョン 2.4 の中にもちに合いそうな項目(例えば食品や農業に関する項目)が無かったが、まず、既存の語彙(コア語彙)のみを用いて共通語彙基盤に合わせた。図 7 はコア語彙のみの共通語彙基盤化である。

第二段階では、共通語彙基盤の中にあるクラス用語一覧(2017年10月時点でバージョン 2.4)の中から、もちデータに合う項目を探した。既存のクラス用語の中で、もちに一番近いと思われる「製品型」を選んだ。製品型は工業製品を想定した用語であるため、郷土食にありそうなプロパティ、例えば食や農業などに関わるで共通語彙基盤化をした。

第三段階では、製品型には無いが、コア語彙の中にある語彙を継ぎ足し、最終的に独自の拡張語彙を作成してもちデータを共通語彙基盤化した。最終的に独自語彙を作成し、共通語彙基盤化をした。図 8 は独自語彙を使用した共通語彙基盤化である。

## (2) LOD(Linked Open Data)

(1) 共通語彙基盤の DMD 作成ツールを使用し、SPARQL エンドポイントで公開することで、もちデータは LOD になった。

LODとは、データを公開して共有するため、コンピュータで処理が出来る形にする技術の総称である。コンピュータ処理が可能なデータの構造化のために、W3C(World Wide Web Consortium)が RDF(Resource Description Framework)というグラフをベースとした基準モデルを作成し、推奨している。RDF については次に説明する。

## (3) RDF(Resource Description Framework)

RDF は、ウェブ上のリソースを記述するための統一された言語体系であり、W3C が 1999 年 2 月に規格化している(W3C:REC-rdf-syntax-19990222)[14]。

従来の HTML では、人間が自ら読み、意味を理解して情報を得た。しかしコンピュータは HTML に記述してある文章の意味を解釈することが出来ない。意味のあるタグ付けとして XML があるが、XML タグは各自が独自の名称を付けるため、独自の名称の意味はコンピュータには理解できない。RDF は記述が統一されているため、SPARQL のような特定の言語を用いることで、機械的に操作することが出来る。

## (4) SPARQL(SPARQL Protocol and RDF Query Language)

SPARQL[15]は W3C が 2008 年にリリースした RDF クエリ言語であり、RDF で記述された XML や Turtle など

の RDF データから検索や抽出をすることが出来る。また、RDF データに SPARQL クエリで問い合わせるための、公開インターフェースとして SPARQL エンドポイントがある。SPARQL エンドポイントには RDF ストアに格納された RDF データを置くことで、SPARQL から問い合わせることが出来る。

共通語彙基盤の DMD を使用することによって、RDF データに変換したもちデータは、代表的な RDF ストアである Virtuoso[16]に格納し、公開 SPARQL エンドポイントに置いてインターネット上に公開した。

## (5) DBpedia

DBpedia[17]には Wikipedia の「インフォボックス (infobox)」と呼ばれる構造化情報(テーブル、カテゴリ情報、および画像、地理座標、外部ウェブページへのリンクなど)を抽出して作られた、問い合わせ可能な統一データセット(RDF 形式にしたデータ)が保存されている。利用者は、DBpedia から SPARQL を使用することでデータセットの情報を検索できる。

## 4. システム構成

従来のシステム(バージョン 1)では、ブラウザ側では Leaflet と OpenStreetMap を使い、Geojson 形式のもちデータを地図上に可視化していた。バージョン 2 では、サーバー側で MySQL と PHP を使用した。今回のバージョン 3 では、REACT[18]と Google Maps API[19]を用い、RDF 形式のもちデータを、Google Maps 上に可視化している。RDF ストアは Virtuoso を使い、SPARQL Endpoint を公開した。図 9 はバージョン 3 のシステム構成図である。

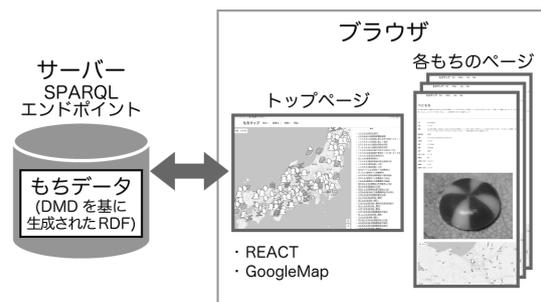


図 9 バージョン 3 のシステム構成図

## 5. 考察

従来のもちマップでは、システムに関しては人が見て分かりやすいもの、データに関しては独自項目を考えて作成していた。今回もちデータを LOD にしたことで、以下の2点の可能性が考えられるようになった。

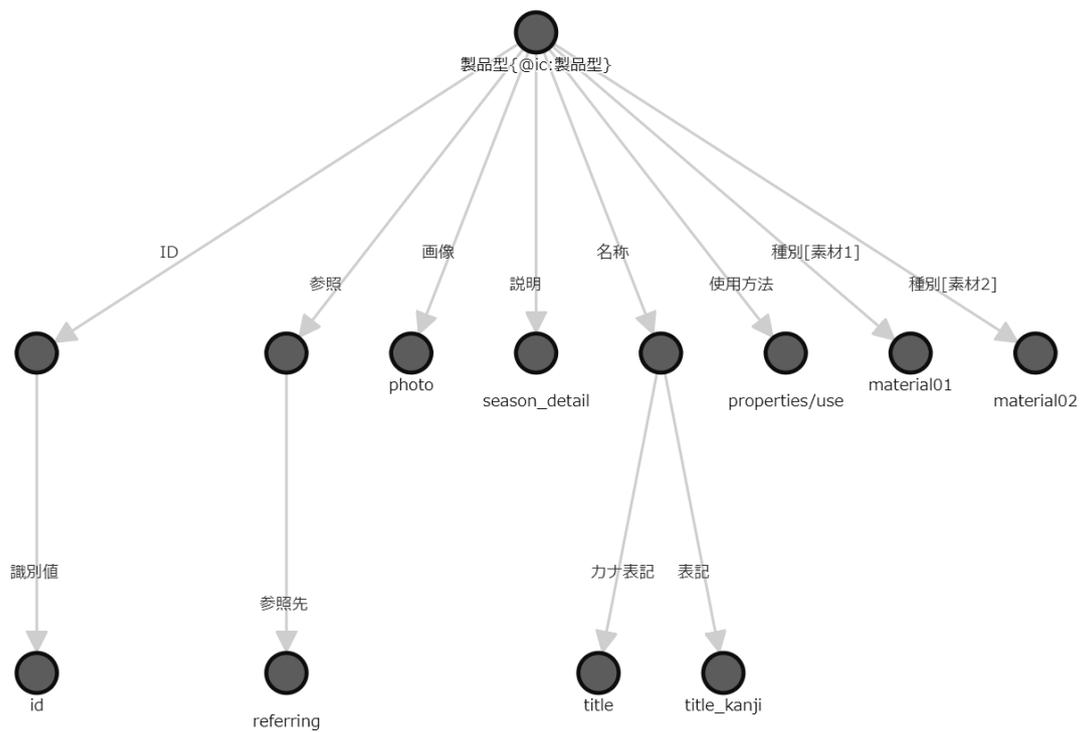


図7 コア語彙のみの共通語彙基盤化

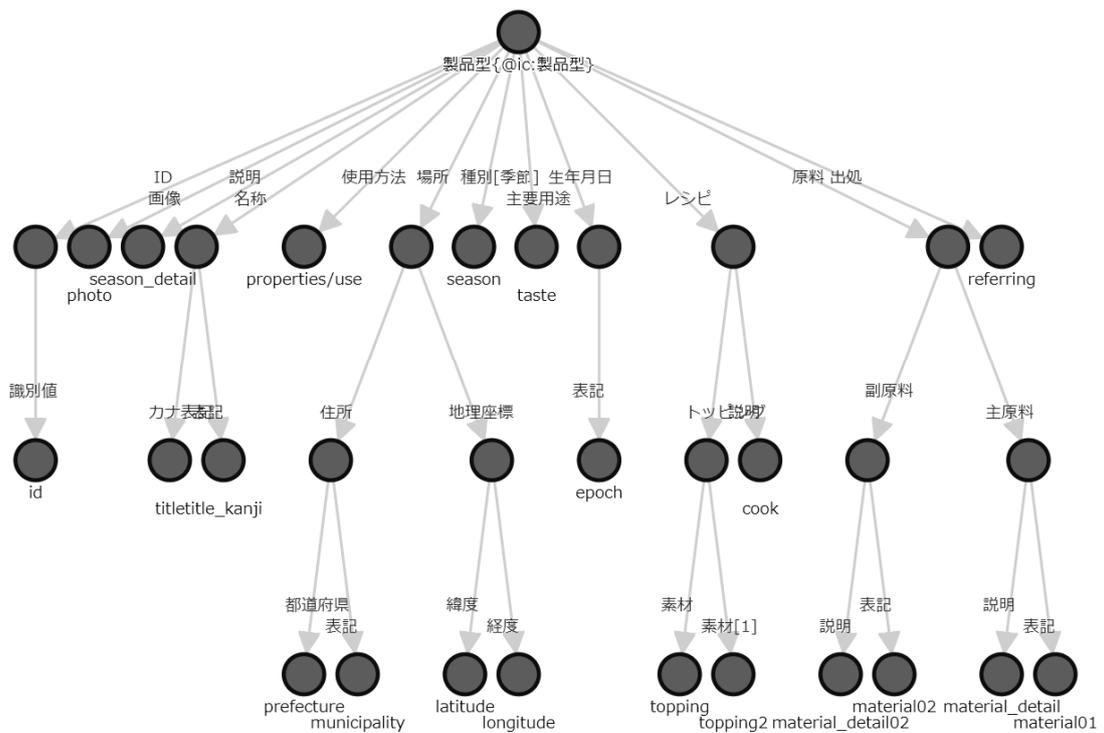


図8 独自語彙を使用した共通語彙基盤化



まれ、データの相互運用性が向上した。また UI 側の改善点としては、各もちの詳細ページが閲覧出来るようになったことがある。これらの改善に伴い、バージョン 3 でも新しい課題が発生した。

- データの構造の改善
- 新しいシステムの UI の改善と、システム自体の検討と改善
- 他の LOD データとの連携

データの構造の改善に関しては、どのようなデータ構造にしたら、元々本研究が目的としている複雑な郷土食の情報を分かりやすく表せるかということを考えなければならない。また、本システムはインターネット上に公開はしているが、まだ広めるための活動や評価をしておらず、独自に考察をしたのみであるため、近いうちにユーザーテストを行う予定である(2018年2月8日現在)。

それに加え、もちマップバージョン 1,2 からの課題である、より利用しやすく、分かりやすい UI への改善や、地域区分や統計情報の工夫、および入力情報の収集・管理方法の検討なども残っている。今後は、今までの課題の解決を進めながら、今回新たに利用した技術を用いてもちマップを改善して行きたいと考えている。

## 謝辞

本論文作成にあたり、名古屋工業大学大学院 特任研究員の河瀬諭先生、来栖川電算の森俊也さん、および実践！共通語彙基盤ワークショップ・ハッカソンの皆様ならび多くの方に大変お世話になった事に深く感謝を申し上げます。

## 参考文献

- [1] 河村 郁江. 郷土食による地域理解支援システム「もちマップ」の試作. 研究報告人文科学とコンピュータ (CH), 2017-CH-113(4), 1-2, (2017-01-28).
- [2] 河村 郁江. “2016 年度版もちマップ”. <http://phirip.com/t/mochimap/e/index.html>, (参照 2018-02-08).
- [3] 共通語彙基盤：IPA 独立行政法人 情報処理推進機構. 実践！共通語彙基盤ワークショップ・ハッカソン. <https://peatix.com/event/298493?lang=ja>, (参照 2018-02-08).
- [4] 奥村 彪生 (解説). 聞き書・ふるさとの家庭料理 第 5 巻 もち・雑煮, 農山漁村文化協会 編集, 農山漁村文化協会, 2002-12.
- [5] 奥村 彪生 (解説). 聞き書・ふるさとの家庭料理 第 7 巻 まんじゅう おやき おはぎ, 農山漁村文化協会 編集, 農山漁村文化協会, 2003-2-25.
- [6] 河村 郁江. “2017 年度版もちマップ”. <http://phirip.com/mochimap/e/>, (参照 2018-02-08).
- [7] 大学共同利用機関法人人間文化研究機構 国立歴史民俗博物館. 民俗語彙データベース. <https://www.rekihaku.ac.jp/doc/gaiyou/goi.html>, (参照 2018-02-08).
- [8] 太田泰弘. 食のターミノロジー(＜小特集＞ドクメンテーションとターミノロジー) ドクメンテーション研究 Vol. 32 (1982) No. 11 p. 555-559, 一般社団法人 情報科学技術協会.
- [9] Google LLC. Google Scholar. <https://scholar.google.co.jp/>, (参照 2018-02-08).
- [10] The Dublin Core Metadata Initiative (DCMI). <http://dublincore.org/>, (参照 2018-02-08).
- [11] Schema.org. <http://schema.org/CreativeWork/>, (参照 2018-02-08).
- [12] 河村 郁江. “2017 年度版もちマップ入力フォーム”. [http://phirip.com/mochi/photo\\_up04\\_04\\_o/](http://phirip.com/mochi/photo_up04_04_o/), (参照 2018-02-08).
- [13] 共通語彙基盤：IPA 独立行政法人 情報処理推進機構. <http://phirip.com/mochimap/e/>, (参照 2018-02-08).
- [14] World Wide Web Consortium. Resource Description Framework (RDF) Model and Syntax Specification. <https://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>, (参照 2018-02-08).
- [15] SPARQL 1.1 Overview. <https://www.w3.org/TR/sparql11-overview/>, (参照 2018-02-08).
- [16] OpenLink Software. Virtuoso. <https://virtuoso.openlinksw.com/>, (参照 2018-02-08).
- [17] DBpedia Japanese. <http://ja.dbpedia.org/>, (参照 2018-02-08).
- [18] REACT. <https://reactjs.org/>, (参照 2018-02-08).
- [19] Google LLC. Google Maps API. <https://developers.google.com/maps/?hl=ja>, (参照 2018-02-08).
- [20] 総務省統計局. e-Stat 統計 LOD. <http://data.e-stat.go.jp/lodw/>, (参照 2018-02-08).
- [21] 経済産業省,内閣官房. RESAS 地域経済分析システム. <https://resas.go.jp>, (参照 2018-02-08).
- [22] Mike Bostock. D3.js. <https://d3js.org/>, (参照 2018-02-08).
- [23] Toshiaki Katayama. <http://biohackathon.org/d3sparql/>, (参照 2018-02-08).
- [24] 加藤 文彦 (著)、川島 秀一 (著)、岡別府 陽子 (著)、山本 泰智 (著)、片山 俊明 (著). オープンデータ時代の標準 Web API SPARQL, インプレス R&D[NextPublishing], 2015-11-13.

# 『本草和名』データベースの構築に向けて Toward Constructing the Database of Honzowamyo

武 倩

劉 冠偉

Wu Qian

Liu Guanwei

北海道大学 文学研究科, 札幌市北区北 10 条西 7 丁目

Graduate School of Letters, Hokkaido University, N10, W7, Kita-ku, Sapporo

**概要:**『本草和名』(918年)は、深根輔仁によって編纂された日本現存最古の本草書である。本書は平安時代を代表する古辞書の『和名類聚抄』に引用されていることから、国語学研究の重要資料となっているほか、その引用典籍には中国・日本の逸書を数多く含んでおり、日中典籍交流史、古医学史の分野でも重要な資料である。『本草和名』の国語学的研究をするには、日中典籍交流史、古医学史に関する知識も必要となる。それらを踏まえた研究を遂行するためにはデータベースの構築が必要とされる。本発表では『本草和名』データベースの構築に向けて、資料の特徴やデータベース化の課題について論じる。

**Abstract:** Honzowamyo written by Sukehito Hukane is the oldest herbal medicine book existed in Japan. It is a very important material for classical Japanese studies as it is cited by Wamyoruijusho, which is one of the representative old dictionaries in the Heian period. It is also important for studying the history of classics exchange between China and Japan and the history of ancient medical, as the book quotes many important documents from China and Japan which have been lost. An interdisciplinary knowledge is needed to study Honzowamyo from the perspective of Japanese language studies. A database is necessary for further studies. In this article, we discuss the characteristics of Honzowamyo and how to build up a database for it.

**キーワード:** 和名類聚抄, 本草書, マルチデバイス

**Keywords:** Wamyoruijusho, herbal medicine book, multi-device

## 1. 本草和名の概要

### 1.1 本草書として

「本草」とは東洋医学で薬の原料となる薬用植物のことで、広く動物・鉱物も含むが、その学問は本草学と呼ばれる。

中国において、本草書の編纂は後漢の『神農本草経』まで遡る。それを校訂・加注したのが陶洪景(六朝)の『神農本草経集注』(『集注本草』『本草集注』とも)である。さらに、この『神農本草経集注』を増補・加注したのは蘇敬(唐)の『新修本草』である。『新修本草』は659年に、高宗の勅命を奉じて編纂されたもので、いわゆる中国最古の勅撰本草書である。

それが奈良時代に日本に輸入され、医学生の教科書などとして重宝された。延喜年間(918年頃)に、深根輔仁が醍醐天皇の勅命を奉じて、唐の『新修本草』

を範にとり、『本草和名』を編纂した。それが日本現存最古の本草書である。

本書は、『新修本草』をはじめとする多くの漢籍から引用している。しかも、その多くはすでに散逸したので、文献の再建においての価値が高い。また、単なる漢籍の抄録に留まらず、薬物の漢名に万葉仮名和訓や日本での産出状況も注記しており、平安時代の語彙・文化・物産などを知る貴重な資料として知られる。さらに、後に日本で編纂された分類体漢和辞書の『和名類聚抄』(934頃)や医学全書の『医心方』(984年頃)等に相当な影響を与えており、辞書や医書の編纂史の研究にも貢献できる。

日本の本草学は、江戸時代には博物学の研究に発展し、明治時代以降は植物学・生薬学に受け継がれる。

## 1.2 諸本

本書が現代に伝わるまでの道のりは順調ではなかった。江戸幕府の医師多紀元簡によって紅葉山文庫で発見されるまで、長く所在を失っていたのである。元簡は本書を校訂し、頭註を加え、序文と提要を附し、寛政八年に版行したことによって世に広めた。しかし、元となる紅葉山文庫の古写本は明治以降再び所在不明となった。

『日本古典全集』所収の、森立之父子が書き入れをした寛政八年版本は入手しやすいため、研究材料として多用されてきた。武(2013)、(2016)は、従来国語学の分野では知られていなかった「万延元年影写本」と「松本書屋本」を紹介した。研究の底本を専ら『日本古典全集』所収の複製本に頼る状況を変えた。

「万延元年影写本」江戸時代に紅葉山文庫で発見された古鈔本を影写したもので、現在台北故宫博物院と無窮会専門図書館神習文庫に所蔵されている。「松本書屋本」(『松本書屋貴書叢刊』所収)は森立之父子らが書き入れをした版本のオールカラー影印版である[12]。

『本草和名』データベースの構築に際し、「松本書屋本」版本と、無窮会専門図書館神習文庫所蔵の「万延元年影写本」を底本にする。

## 1.3 体裁及び版本の考注

本書は、唐の『新修本草』の編次に従って、千種以上の薬物を収録している。体裁は、漢名を見出し語に、下に異名、和訓、産地などを記す。薬の効能・配伍・採取時節などは省かされており、漢籍本草書や医書を読むための手引きとしての性格を併せ持つといえる。

版本には、本文内容に加えて、多紀元簡や森立之父子ら江戸後期・明治の考証学者による校訂も見受けられる。その校訂に版本と古鈔本との相違に関連するものがある。

例えば、「雲母」條(図2)では、「一名地涿」の「涿」字に対して、多紀元簡は上欄で「按涿原本缺今据証類校補」という頭註を加えている。この「原本」は古鈔本のこと、古鈔本において「涿」という字が欠けており、元簡は版元を作る際に、『証類本草』によってそれを補ったという。古鈔本を影写した「万延元年影写本」(図1)を見ると、「涿」が空白となっていることが確認できる。

前掲の例では、「原本」脱字を示しているが、その他に、両本の間で衍字、異体字などの相違を示すものもある。今回のデータベースに反映させるのは脱字と衍字に止まり、異体字を今後の課題としたい。

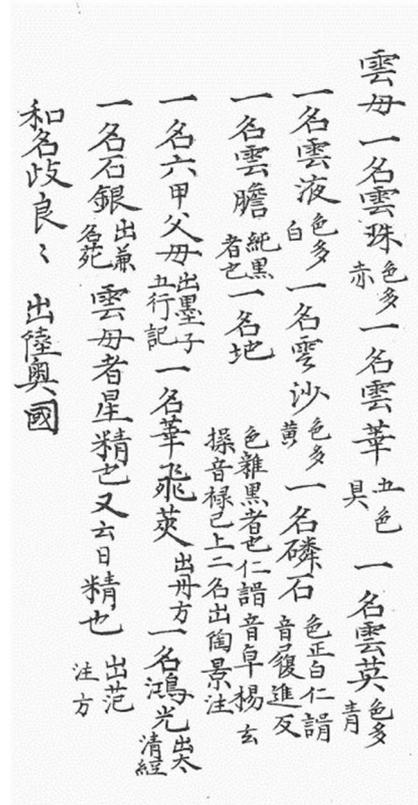


図1 無窮会蔵万延元年影写本

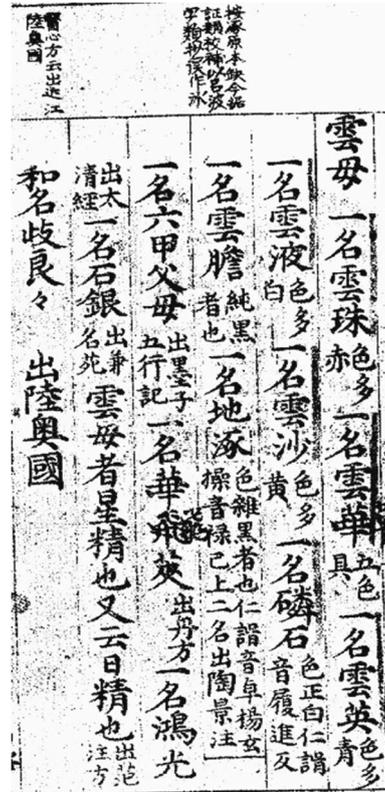


図2 日本古典全集所収版本

## 1.4 和名

『本草和名』の和名は万葉仮名で記されている。本書に使用される万葉仮名を表1にまとめる。

表1 『本草和名』万葉仮名字母

ア	イ	ウ	エ <sub>1</sub>	オ
阿	以 伊	宇	衣	於
カ	キ	ク	ケ	コ
加 迦	岐 支 木	久	介	古 己
サ	シ	ス	セ	ソ
佐	之	須	世	曾
タ	チ	ツ	テ	ト
多 太	知	都	天	止
ナ	ニ	ヌ	ネ	ノ
奈	尔	奴	祢	乃
ハ	ヒ	フ	ヘ	ホ
波 巴	比	布	倍	保
マ	ミ	ム	メ	モ
末 万	美 三	无 牟	女	毛
ヤ		ユ	エ <sub>2</sub>	ヨ
也		由	衣 江	与
ラ	リ	ル	レ	ロ
良	利	留	礼	呂
ワ	ヰ		エ	ヱ
和	為 委		恵	乎

データベースで仮名検索できるように、データ入力の際に、万葉仮名を平仮名に直して情報として付け加える。

## 2. データベース構築の目的

### 2.1 和名の漢字表記についての研究

『本草和名』は1034種の薬物を収録しており、その中の594種に和名が付されている。体裁の紹介で述

べたように、本書は漢籍本草書や医書を読むための手引きとして用いられていたと考えられる。薬の効能・配伍・採取時節などを示すよりは、薬の同定に重点を置き、可能な限り多くの異名を列挙する傾向がある。例えば、「雲母」の例では、「雲珠」「雲華」「雲英」「雲液」「雲沙」「磷石」「雲膽」「地涿」「六甲父母」「華飛莢」「鴻光」「石銀」「星精」「日精」の14個の異名も挙げている。

また、一つの見出し漢名に一つの和名しか付されているとは限らず、二つの和名を持つのは95項目で、三つの和名を持つのは8項目である。

さらに、項目間の関係を見ると、例1のように、同じ漢名が複数の項目に出現する場合がある。

例1:「雲母」

雲母 一名雲珠【色多赤】…和名岐良々 出陸奥国（上三03オ3）<sup>1</sup>

落石 【蘇敬注云苞絡石木而生故以名耳】…一名雲母一名雲華一名雲珠一名雲英【已上五名出積薬性】…和名都多（上七20表6）

雲實 一名員實一名雲英一名天豆【蘇敬注黄黒似豆故以名之】一名雲母【苗名也出蘇敬注】和名波末佐々介（上七23オ4）

これとは別に例2~5のように、同一和名が複数の項目に出現する場合（次に示す）もある。

• 二項目に出現：計35語

例2:「乎止乎止之」

赤箭 【蘇敬云遠看如箭】一名離母一名鬼督郵【仁譚音尤】一名神草一名獨揺一名當苦一…和名乎止乎止之一名加美乃也（上六11ウ9）

鬼督郵 一名獨揺草一名餘毒草【出范注方】一名悶狗【出積薬又積薬性悶作悶字】和名乎止乎止之（上七25ウ1）

例3:「於毛多加」

澤蔞 【仁譚音昔楊玄操音私也反】一名水寫【一名及寫】一名芒芋一名箭寫…和名奈末為一名於毛多加（上六14オ6）

鳥芋 一名籊姑…一名茨菰【澤泻之類也已上出蘇敬注】鳥苳【出崔禹】…和名於毛多加一名久呂久和為（下十七31ウ1）

（後略）

<sup>1</sup>「上三03オ3」は『本草和名』版本の「上冊第三卷3丁表3行目」を表す。「【】」で括る部分は小字双行です。以下同じ。

- 三項目に出現：計 2 語

例 4：「阿末奈」

黄精 一名重楼【楊玄操音直竜反】…一名飛英

【花名】一名流精【根名】一名馬箭…和名  
阿末奈一名也末恵美（上六 13 オ 4）

麻黄 一名龍沙一名卑相一名卑塩一名狗骨【出  
積薬性】和名加都弥久佐一名阿末奈（上八  
26 ウ 8）

白薇 一名白幕一名薇草一名春草一名骨美一名  
白蜜曹【出積薬性】…和名美奈之古久佐一  
名久呂女久佐一名阿末奈（上八 30 ウ 9）

例 5：「以奴衣」

蘇 野蘇【不香似荏者】一名桂荏一名卿【已上  
二名出兼名苑】和名以奴衣一名乃良衣（下  
十八 37 ウ 7）

假蘇 …一名薑芥一名荊芥【蘇敬注云薑者荊声  
訛也是菜中荊芥是也】和名乃々衣一名以奴  
衣（下十八 38 オ 3）

香薷 【楊玄操音而由反】一名胡薷【本名胡薷  
勒諱胡故名香薷出兼名苑】…和名以奴衣一  
名以奴阿良々岐（下十八 38 オ 6）

このように、『本草和名』においては、「見出し漢名⇔異名⇔和名」の対応関係が複雑な様相を呈している。国語学の観点からは、本書が専門用語を集めた辞書一葉名辞書として位置づけられる。

本書における和名の漢字表記を詳しく考察するためには、『本草和名』データベースの構築が必要である。

## 2.2 辞書史の研究

古辞書においては、『本草和名』は、『和名類聚抄』の序文にも名を載せられ、『和名類聚抄』の重要な出典の一つであると知られている。

つとに、『本草和名』を国語資料としての全貌を描いた築島(1965)の研究がある。その後、『本草和名』と『和名類聚抄』との引用関係については、河野(1983)、河野(1988)、宮澤(1997)、宮澤(2010)の研究が出た。『医心方』と『本草和名』『和名類聚抄』の関係については、松本(1980)、築島(1987)、加藤(2005)、松本(2007)の考察がなされた。

また、観智院本『類聚名義抄』との引用関係について調査した研究に河野(1992)、図書寮本『類聚名義抄』との関連に調べた研究に呉(2000)が挙げられる。

今後、『本草和名』データベースとその他の古辞書データベースとの連携によって、前述の先行研究をさらに進展させることを願っている。

## 3. データベース構築とインターフェースの開発

### 3.1 入力方針

前述のように、底本には「松本書屋本」版本と、無窮会専門図書館神習文庫所蔵の「万延元年影写本」を用いる。実際のデータ入力には、版本をベースとして、影写本と相違のある箇所をマークアップし、注を附する。データを入力するソフトウェアには Excel を使用する。データベースの範囲は掲出項目、本文、および元簡による頭注のみで、序文などは含まない。

現在の進捗状況は、上冊の入力が終わっており、下冊の入力が半分ほど進んでいる。以下で行われるデータの分析は、上冊のみに止まり、下冊の入力が終了次第、稿を改めて報告したい。

本書にある文字はユニコードによる符号化を行い、符号化が難しい漢字は「」で括って、漢字構成記述文字列(IDS)で記入する。結果として、表 2 の示すように、上冊においては、ほぼ全ての見出しは符号化ができています。本文も表 3 の示すように、拡張漢字 B (Ext. B) まで 99%(延べ)符号化でき、IDS を利用すれば全ての本文を入力することができる。

本文にある割注は「【】」で、「一名」の後ろにある別名は「<sup>1</sup>」で、松本本の上欄にある頭注に対応する漢字は「〈〉」で括る。

表 2 見出し符号化率

ブロック	延べ数	異なり数
CJK	934(98.6%)	497(97.3%)
Ext. A	8(0.8%)	6(1.2%)
Ext. B	5(0.5%)	5(1.0%)
総計	943(99.7%)	508(99.4%)

表 3 本文の符号化率

ブロック	延べ数	異なり数
CJK	18,714(98.9%)	1,627(92.1%)
Ext. A	16(0.1%)	10(0.6%)
Ext. B	32(0.2%)	23(1.3%)
踊り字	45(0.2%)	1(0.1%)
総計	18,807(99.4%)	1,661(94.0%)

### 3.2 フィールドの構成

本データベースは一つのテーブルからなる。テーブルは id、book、vol、page、line、entry、def、diff-type、

diff-def、japanese-name、japanese-reading の 11 個のフィールドで構成した。図 1 として示した「雲母」を例にして、各フィールドの内容と例文は表 4 に示す。

その中、japanese-reading フィールドは前述の 1.4 節に従って現代仮名遣で翻刻したものである。両本相違の diff-type と頭注の「diff-def」は本文の def にある「〈〉」で括られたものに対応する。「〈〉」で括られたものが複数ある場合、「;」で区切り、出現順に入力する。

表 4 本草和名データベースのフィールド構成

フィールド	内容	例
id	出現順 ID	HW0010
book	冊	上
vol	巻	三
page	掲出頁	03表
line	掲出行	3
entry	見出し	雲母
def	本文	一名「雲珠」【色多赤】 一名「雲華」【五色具】 一名「雲英」【色多青】 一名「雲液」【色多白】 一名「雲沙」【色多黄】 一名「磷石」【色正白 仁譚音履進及】一名 「雲膽」【純黒者也】一 名「地〈涿〉」【色雜黒 者也仁譚卓楊玄操音 祿已上二名出陶景 注】一名「六甲父母」 【出墨子五行記】一名 「華飛莢」【出丹方】一 名「鴻光」【出太清經】 一名「石銀」【出兼名 苑】雲母者「星精」也 又云「日精」也【出範 注方】和名岐良々出 陸奥国
diff-type	両本相違	空字
diff-def	頭注	按涿原本缺今据証類 校補
japanese-name	和名	岐良々
japanese-reading	現代仮名遣 による読み	きらら

### 3.3 インターフェースの実装

本データベースのインターフェースは図 3 の左の検索欄と右の結果欄からなる。

検索欄では利用度の高い「見出し」、「本文」、「和名（ひらがな）」の三つの条件で検索できる。検索した結果は右に表示される。結果の見出しは大きいサイズの文字、和名と本文は小さいサイズの文字で表示される。図 3 は和訓を「あ」という条件で検索した画面である。

インターフェースはマルチデバイス対応済みであり、PC 以外の端末でも利用しやすいレイアウトにしている。図 4 はスマートフォンでの利用画面である。



図 3 インターフェースの検索画面



図 4 スマートフォンでの利用画面

これまで、古辞書データベースの検索インタフェースのマルチデバイス対応に関しては、池田(2014)が推進する平安時代漢字字書総合データベース(HDIC)に対応させた HDIC Viewer の開発を手始めに、夢梅本『倭玉篇』全文データベース検索システムの試作も行なっている[7][16][17]。今回の『本草和名』データベースの開発では、それらとベースとなるシステムは同一として、特にインタフェースの改善に力を注いだ。

#### 4. 付記

本研究は武田科学振興財団「杏雨書屋研究奨励」の支援のもとで実施されている。

平素は、池田証壽先生から、丁寧かつ熱心なご指導を頂いており、ここに感謝の意を表します。

また、資料を提供して頂いた無窮会専門図書館の関係各位に感謝を申し上げる。

#### 参考文献

- [1] 池田証壽, 平安時代漢字字書総合データベースの構築, 北海道大学文学研究科紀要, 2014, vol. 142, p. 79-90.
- [2] 呉美寧, 図書寮本類聚名義抄所収の植物一和名類聚抄・本草和名との関連を中心に, 古辞書と J I S 漢字, 2000, vol.3, p. 29-38.
- [3] 加藤大鶴, 『医心方』字音注記出典と加點方針についての一考察: 『本草和名』『和名類聚抄』との比較を通じて, 論集, 2005, vol. 1, p.145-165.
- [4] 河野敏宏, 『和名類聚抄』と『輔仁本草』の関係について: 『和名類聚抄』漢文本文に関して, 岡大國文論稿, 1983, vol. 11, p.16-26.
- [5] 河野敏宏, 『和名類聚抄』の音注の文献的性格: 『本草和名』の音注との比較による, 愛知学院大学論叢(一般教育研究), 1988, vol.35(3・4), p.49-74.
- [6] 河野敏宏, 観智院本『類聚名義抄』所引の『本草和名』について, 日本語論究, 1992, vol. 2, p.69-85.
- [7] 高橋大希, 劉冠偉, 池田証壽, 夢梅本『倭玉篇』全文テキストデータベースの構築, 言語資源活用ワークショップ 2016 発表論文集, 2016, p.49-56.
- [8] 築島裕, 本草和名の和訓について, 国語学研究, 1965, vol. 5, p.1-10.
- [9] 築島裕, 『本草和名』の和訓と『医心方』の万葉仮名和訓, 国書逸文研究, 1987, vol. 20, p.3-12.
- [10] 武倩, 『本草和名』の諸本に関する一考察—万延元年影写本と全集本との関係を中心に—, 訓点語と訓点資料, 2013, vol. 131, p.43-52.
- [11] 武倩, 松本書屋本『本草和名』について, 北海道大学大学院文学研究科研究論集, 2016, vol. 15, p.51-60.
- [12] 松本一男, 松本書屋貴書叢刊, 第 1 卷, 谷口書店, 1993
- [13] 松本光隆, 書陵部蔵医心方・成篋堂文庫蔵医心方における付訓の基盤: 和名類聚抄・本草和名との比較を通して, 鎌倉時代語研究, 1980, vol. 3, p.133-154.
- [14] 松本光隆, 第五節 医心方天養二年点における付訓の基盤—和名類聚抄・本草和名との比較を通して—, 平安鎌倉時代漢文訓詁語史料論, 汲古書院, 2007.
- [15] 宮澤俊雅, 倭名類聚抄諸本の出典について, 北海道大学文学部紀要, 1997, vol.45, no. 2, p.1-29.
- [16] 宮澤俊雅, 倭名類聚抄と漢語抄類, 倭名類聚抄諸本の研究, 勉誠出版, 2010.
- [17] 劉冠偉, 李媛, 池田証壽, 平安時代漢字字書総合データベースの拡張と和訓対応, 研究報告人文科学とコンピュータ, 2015 vol. 2015-CH-106, no. 4, p. 1-8.
- [18] 劉冠偉, 李媛, 池田証壽, 多漢字文献の検索効率向上とマルチデバイス対応の試み, 研究報告人文科学とコンピュータ, 2017, vol.2017-CH-114, no. 6, p. 1-4.

# 方言分布・言語地図データベース — 時空間情報を持つ言語データ —

## Database on Dialectal Distributions and Dialectal Maps: Linguistic Data Including Spatial and Temporal Information

大西拓一郎

Takuichiro Onishi

国立国語研究所 言語変化研究領域, 東京都立川市緑町 10-2

National Institute for Japanese Language, 10-2, Midori-cho, Tachikawa, Tokyo

**概要:** 方言は、土地によることばの異なりである。ことばとしての本質において、標準語との差異はない。ただし、方言である以上、使われる場所についての情報は欠かすことができない。したがって、方言情報には通常の言語情報に加えて、空間情報が必須である。空間情報を緻密にすると、分布が求められ、地図として表現が可能になる。日本の方言学は分布や地図の情報を豊富に持ち、われわれはそのデータベース化を進めている。その現状と展開を報告する。

**Abstract:** Dialects are variations of language decided by space. There are no differences between dialects and standard Japanese in the nature of language; however, dialectal data require spatial information besides linguistic information. Adding enough spatial information on dialectal data clarifies dialectal distributions and makes it possible to express them on maps. Japanese dialectology has rich data of distributions and maps allowing for the creation of a database of these dialectal distributions. In this paper, I report the present and future works with regard to the database.

**キーワード:** 方言、方言分布、方言地図、GIS

**Keywords:** dialect, dialectal distribution, dialectal map, GIS

### 1. はじめに

方言分布と方言地図のデータベースの公開を開始してから、すでに約 15 年になる。その概要(沿革・現状)を報告し、諸分野との連携を展望する。

### 2. 方言地図とは

方言は、歴史的な同系統言語の場所による異なりであり、言語としての本質は標準語と異なる。そのような方言がどのように存在しているのか、つまり分布を視覚化し、地図として表現したのが方言地図である(図 1)。方言地図は、言語地図と呼ばれることもある。

方言地図は、方言という特定の対象を表現することから、主題図であり、地形図のような一般図とは区別される。したがって、方言地図において、どのような表現方法を採用するかには、地形図の地図記号のような一般則はない。

分布の表し方は、語形を地図上に書き込む語形記入法、分布域を色や塗りつぶし方で表現する塗りつぶし法、語形を記号(シンボル)に置き換えて示す記号法の 3 方法がある。語形記入法は、世界最初期の『フランス言語図巻』が採用したが、煩瑣で分布が把握しづらく、地図としての効果が低い。塗りつぶし法は対象

となる方言形式の種類が多いと限界があるとともに複雑に入り組んだ分布に対応しきれない。多くの方言地図は、語形や分布の多様性に対応可能な記号法を採用している。地図要素の一つであるポイントで扱う記号法は、対応するデータの性質(各地点の話者)との親和性も高い。

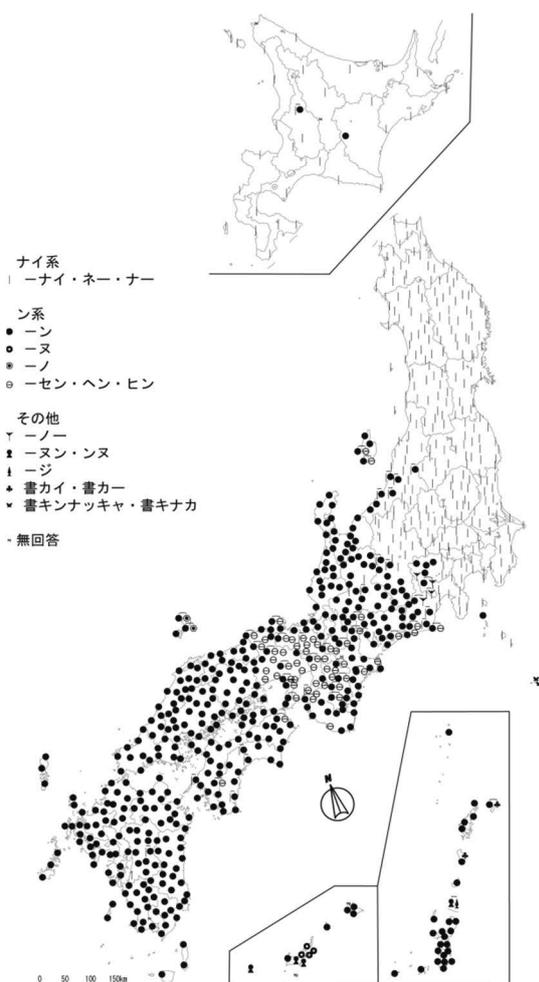


図1 書かない(GAJ2-80に基づく)

### 3. データ収集調査

方言地図のためのデータが、事前に用意されていることはほぼ皆無なので、一般にデータ収集のためのフィールド調査を実施し、それをもとに地図が作成される[1]。各地図集の内容はそれぞれ異なるが、以下に一般的なモデルをあげ、おおまかなイメージを示す。

#### ・対象領域

全国を対象とするような大規模な調査研究は、国立国語研究所が行うことがほとんどである。それに対し、

単一もしくは複数の市町村、あるいは都道府県が対象とされることもあり、河川の流域が選ばれることも少なくない。このような狭域を対象とする場合は、大学の研究室が中心になって実施されることが多い。

#### ・調査項目・方法

何を調べるかであるが、対象地域内に異なり(方言差)が想定される対象を選定するのが基本である。多くの場合は、動植物名などの対象(「かたつむり」「桑の実」等)や標準語形式(「行かない」「行かなかった」等)を絵も含め、意味として提示し、それを表す語形を求める。項目数は200程度が多い。

#### ・調査地点

対象地域の広さによるが、地図にする場合、図郭内に最低200地点程度は含まれていないと、見栄えが良くない。市区町村程度を対象にする場合は、基本的に全集落を目標にすることが多い。全国レベルの場合は、やや幅があり、500~2000地点程度を対象に実施されてきた。

#### ・話者

言語は変化する。また、使用者の移動は言語に影響を及ぼす。したがって、話者の条件をある程度そろえなければ、場所による異なり以外の属性が地図上の言語変異に投影されてしまう可能性が否定できなくなる。

そこで、話者の年齢層や居住歴などの条件を一定にすることが必要である。多くの場合は、高齢(例えば、70歳以上)で基本的に生え抜き(もしくは移動期間を特定年数以内に納める)といった条件が設定される。

### 4. 日本の方言地図

日本では、これまで400冊以上(地図枚数30,000枚以上)の方言地図集が作成されてきた。この数字は、世界最高レベルである。

#### ・広域地図

最初に刊行された方言地図集は、国語調査委員会による『音韻分布図』(1905年)と『口語法分布図』(1906年)であり[2]、全国を対象としている。先に挙げた『フラ

『方言地理学の方法』の刊行(1902-1909年)からそれほど遅れてはいない。

日本全国を対象とした方言地図集は、その後、国立国語研究所による『日本言語地図』(LAJ:2400地点)[3]と『方言文法全国地図』(GAJ:807地点)[4]があり、21世紀に入ってからLAJ・GAJとの経年比較をおもな目的とする『新日本言語地図』(NLJ:554地点)[5]が刊行されている。

#### ・狭域地図

1970年代から、地域ごとの方言地図集が盛んに刊行されるようになる。これは方言地図を用いた研究分野＝言語地理学の展開と連動する。柴田武・徳川宗賢・W.A.グロータース・馬瀬良雄による新潟県糸魚川地方の方言分布調査が1957～1961年に実施された。この調査を基盤として、柴田武は『言語地理学の方法』により言語地理学を整備した[6]。それは柳田国男による方言周圏論[7]を近代化し、目的と方法を明確化するものであった。そこでは、地域を対象にした調査、調査結果の地図化、地図に現れた分布の分析など、それぞれの方法が具体的に明示された。

『言語地理学の方法』の刊行と同時期にLAJの刊行が開始された。これが日本の方言学の中での方言分布研究・言語地理学の隆盛を促し、狭域を中心に多数の方言地図集ならびに研究文献が生産された。その推移を図2に示した。

## 5. 方言地図の作成と方言分布データ

LAJとGAJの場合、調査時は調査票に回答を記載し、そこから各地点の各項目を1枚ずつ転記したカードが、国立国語研究所に報告された。調査票は調査者の手元に残され、国立国語研究所は報告されたカードを保管する。すなわち、調査結果は二重化されることになる。

この背景には、国語調査委員会の苦い経験がある。『音韻分布図』『口語法分布図』の刊行後、第二次全国調査が実施され、その結果は当時の文部省に送付・保管されていたが、関東大震災ですべて焼失した。そこで、このような二重化により、完全消失の危機回避をはかった。

方言分布の調査データは、そのままでは地図にはな

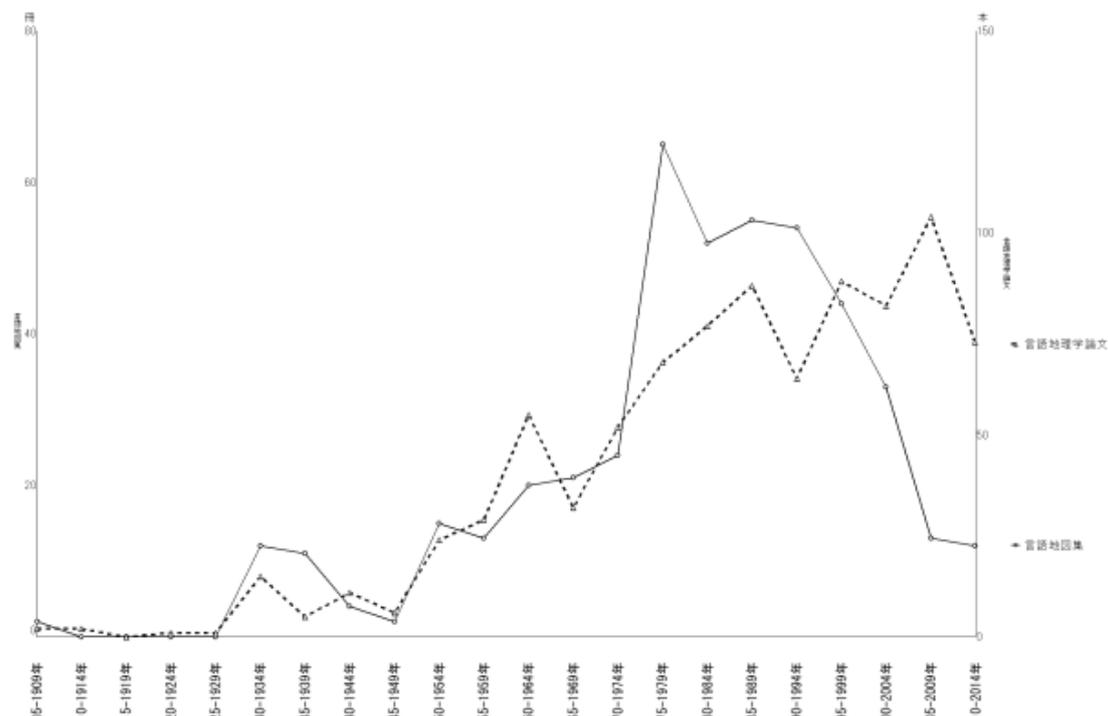


図2 方言地図集の刊行と方言地理学研究文献数の推移[8]

らない。報告された回答の中には、地図として、採用されないものもある。転記ミスによるエラー(例:項目 A の回答なのに B の回答として転記)、誤答(例:動詞否定過去形を回答すべきところを動詞否定現在形で回答)などがその例である。また、採用すべきところを抽出したり(例:動詞否定過去形として報告されたアイツワイカナンダゾからイカナンダを抜き出す)、各地図のねらいに応じて、地図見出しに統合する(例:動詞否定過去形がねらいなので、イカネカッタ・エカネカッタ・イガネカッタ・エガネカッタ・イガネガッタ・エガネガッタは、ひとつの地図見出し「ネカッタ」にまとめる)等の過程が必要である。

この作業を経て、各地図で扱う地図見出しの方言語形が決まる。ここで確定した地図見出しと地図記号の対応が凡例である。

LAJとGAJでは、地図見出しをもとにカードを整理・分類し、LAJとGAJの第4集まではこのカードをもとに手作業で地図化が行われた。地図見出しと地図上の記号の対応を示す凡例に基づき、白地図上に地図記号を専用のハンコで押印する作業である。これを原稿として、印刷所に入稿した。

GAJの第5集以降は、コンピュータによる地図作成に移行した。GAJでは、当初より、報告された全カードのデータ入力を行い、資料一覧として刊行(GAJ付録の解説書に収録)・公開していた。

このデータも活用することで、調査地点・回答・地図見出しの一覧表が作成できる。調査地点は、空間上の場所を指定する空間データであり、回答・地図見出しといった言語(方言)情報は、それぞれの空間データの属性情報である。つまり、この一覧表は、地理行列にはほかならない。これに地図見出しと地図記号の対応表をリンクするなら、空間データと地図記号がつながることになる。このようにして、方言地図はコンピュータで作図可能になった。

GAJは全6巻刊行後、第4集以前にもさかのぼって、コンピュータ作図を行い、エラーを修正した上で、画像ファイル(PDF)の公開を行っている。

NLJのための調査(FPJD)は、2010~2015年に実施した。GAJの作業実績を基盤として、当初から、データベース化とコンピュータ作図を想定して計画を進めた。その結果、調査終了の翌年には方言地図集として、調査結果を公開できた。

## 6. 方言分布データベース

GAJは、調査地点・回答・地図見出しの全データ388,659件を各集に分けて(第1集:59,042件、第2-3集:89,231件、第4集:58,878件、第5集:83,574件、第6集:97,934件)、エクセルファイルで公開している。当初はシフトJISを用いたテキストファイルで公開していたため、音声記号はアスキー文字の組み合わせで表現することで対応していたが、現在はユニコードで音声記号を表現している。

NLJは、FPJDのデータベースを元にするものであり、FPJDのデータベースにNLJの地図見出しデータを付与し、エクセルファイルで公開している(全データ:145,929件)。

なお、LAJについては、別途、国立国語研究所の熊谷康雄氏がデータベース化を進めている[9]。

## 7. 方言地図データベース

### ・目録

前述したように400冊超の地図集が刊行されており、その書誌情報の把握には、データベース化が必須である。代表的なものにしぼった目録が出されることもあった[10]。刊行地図の全体は、国立国語研究所図書館に所収されているもの(日本の方言学においては、刊行物を国立国語研究所図書館に寄贈する慣習があり、日本の方言学の学会である日本方言研究会ではそれをもとに文献リストを作成している[11])をもとに大西が言語地理学の動向を探る[12]作業の過程で集約した。その内容は、当初、日本方言研究会のサイトで公開し、その後、主要なものを選定して目録化した[13]。

このような地図集の目録整備に続き、各地図集に収録された30,000枚以上の地図項目をデータベース化

した。このデータベースは、FPJDを実施する際に、調査項目作成の基盤とすることをねらって作成された。そのため、各地図のキーワード的な分類指標が含まれており、それをもとに地図を検索することも可能である。

以上の地図集書誌目録、地図目録は、FPJD実施途上の2014年2月に公開した。

・画像

目録を手がかりにすることで、原図の存在が確認できる。もとの地図集が手元がない場合は、国立国語研究所図書館(研究図書室)に足を運べば、原図にたどり着くことができる。問題は、狭域地図集の多くが研究室を刊行元とする私家版で、国立国語研究所まで行かないと見られないことである。

そこで、2017年度から、地図を画像にして公開することを開始した。

公開している地図画像には、画像ファイルの中に地図図郭の経度緯度を格納したジオタグ付き画像ファイルが含まれる。ジオタグ付きファイルは、代表的なGISであるArcGIS・SIS・QGISにおいて、地図上の正しい位置に直接表示・オーバーレイできるようになっている。したがって、同じ項目について、異なる地図集の地図をGIS上で表示することにより、対象域を広げる形で分布を確認するようなことも可能である(図3)。

ファイル形式はTIFFを採用している。これは、事前に検討した結果、これらのGISに共通してワールドファイルとして、ジオタグが有効なのがTIFFであったことによる。また、TIFFなので、多少サイズが大きいものの、

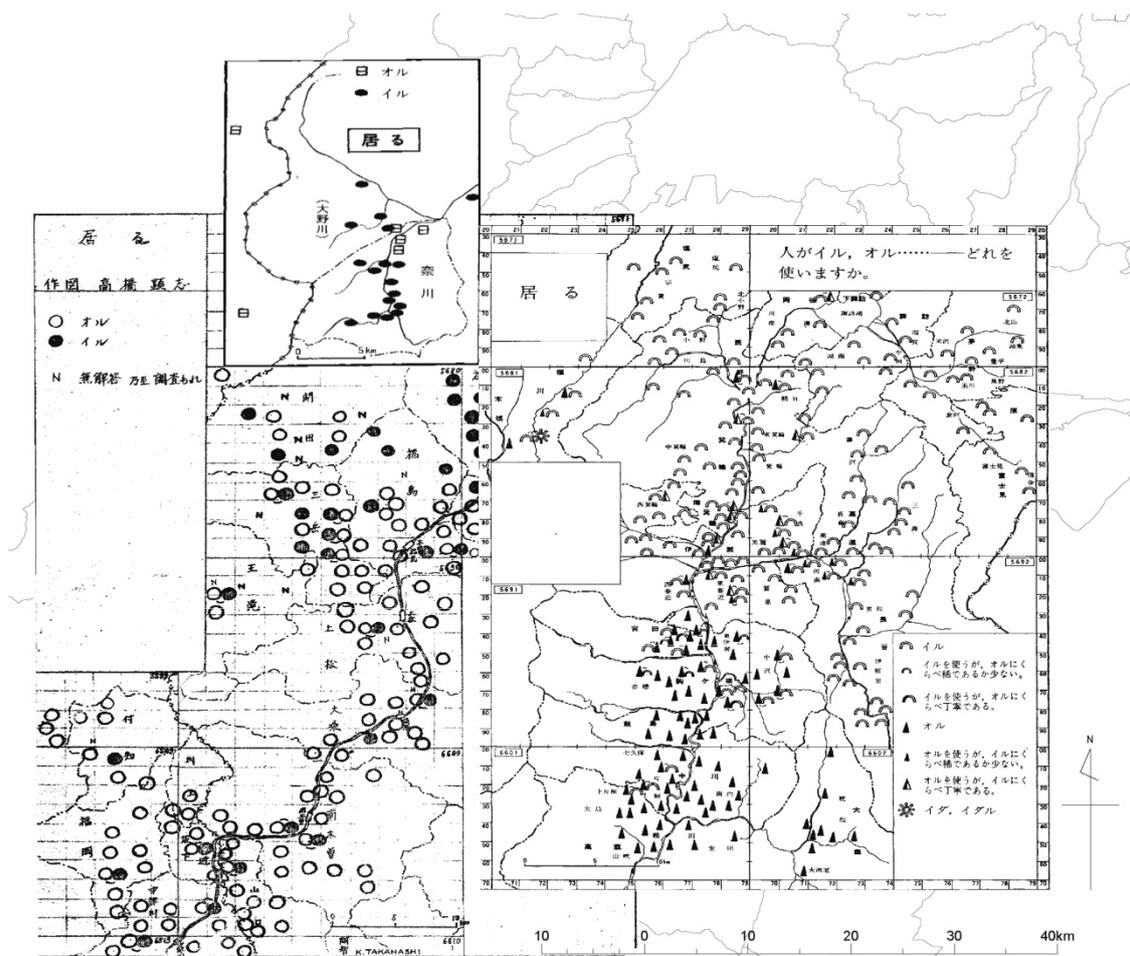


図3 長野県木曾・上伊那・諏訪地方の「いる」

(信州大学方言研究会編 1976『木曾及びその周辺地方の言語地図 語彙篇I』、馬瀬良雄 1980『上伊那の方言』、馬瀬良雄 1997『信飛国境地帯 奈川村の方言』を行政界図上に重ねて表示している。)

単に閲覧する場合も問題は生じない。

なお、初期の狭域地図集は、1960～1970年代に作成されたものが多く、刊行からすでに50年を経ている。用紙やインク類の品質に起因すると考えられる劣化が生じているものも少なくない。画像ファイル化は、資料の保存も目的に含んでいる。

以上のデータベースは、ウェブサイト「方言研究の部屋」[14]から閲覧やダウンロードができる。

## 8. GISでの活用

GAJ・LAJの方言分布データベースは、調査地点と方言情報の組み合わせにより構成されている。このデータベースにおける調査地点は、「地点番号」により6桁の数字でコード化されている。これは、日本全体を、5万分の1地形図が縦横10枚ずつ(計100枚)入るメッシュで縦横10等分し、左上から00、01、02と数字を与え、右下を99とする。そのメッシュの中の地形図に対し、同様に00から99までの数字を与える。さらに地形図の中を縦横10等分して、同じく00から99までの数字を与える。以上の2桁の数字を3回組み合わせることで地点を特定する(地点が収まったメッシュの番号を与える)方法で、国立国語研究所地点番号システムと呼ばれている(図4)。

地形図を基盤にしているため、各図郭の経度緯度は確定でき、それを基にして地点番号が与えられたメッシュの中心の経度緯度を算出することが可能である。すなわち、GAJ・LAJは調査地点データに対し、空間位置を指定する属性を与えることができる。

なお、このシステムは図4の(AとBの枠組みがある)上段から理解されるように、AとBの間で最初の2桁が重複する可能性がある(境界は北緯29度)。実際には、海域と陸(島)域の関係で、ほとんどの場合、重複が回避される(正確には、小笠原で重複が発生するが、LAJ・GAJは小笠原を調査しなかった)。

小地域を対象にした狭域地図は、さらにもう1段メッシュ区切りを追加し、8桁にすることでこの方式を踏襲することが多く、6桁の場合と同様に与えられた地点番

号から経度緯度が求められる。なお、地点番号から経度緯度を求める計算式は、大西のウェブサイト[15]に挙げられている。

NLJでは、最初から調査地点を経緯度の数値の組み合わせでコード化しており、地点の位置指定は、さらに簡便に求められるようになっている。

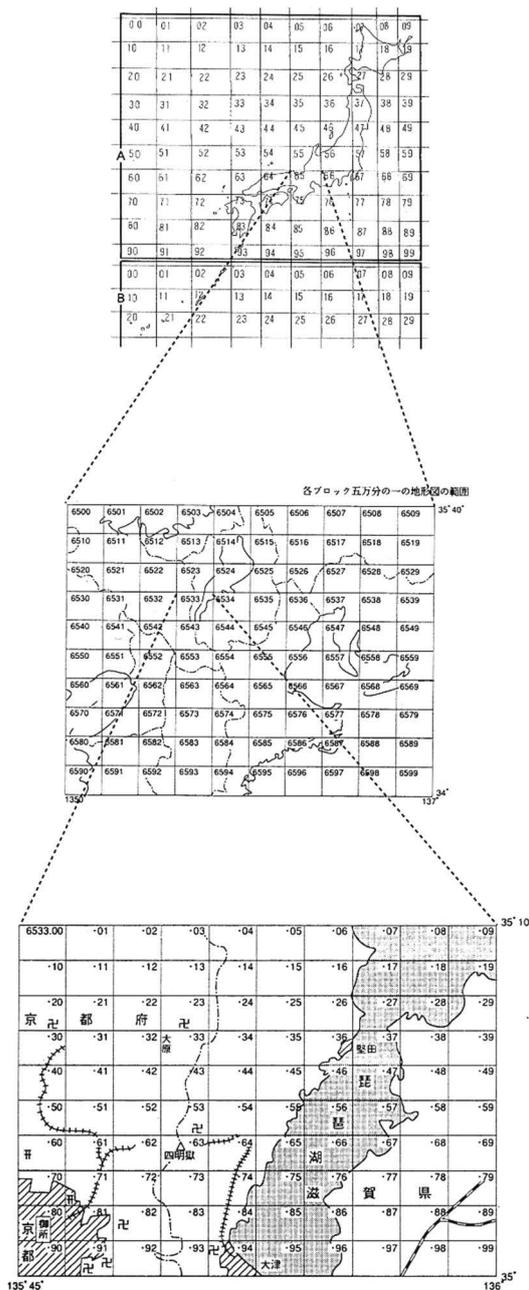


図4 国立国語研究所地点番号システム

以上のように、公開している方言分布データは、GISによる利用の便が高い。

例えば、方言分布と河川や標高を基にした地形との関係を把握することが可能である(図5)。

また、GAJとNLJを較べることで分布の経年変化を把握し、それと人口密度に基づく社会状況との関係を考察することもできる(図6)。

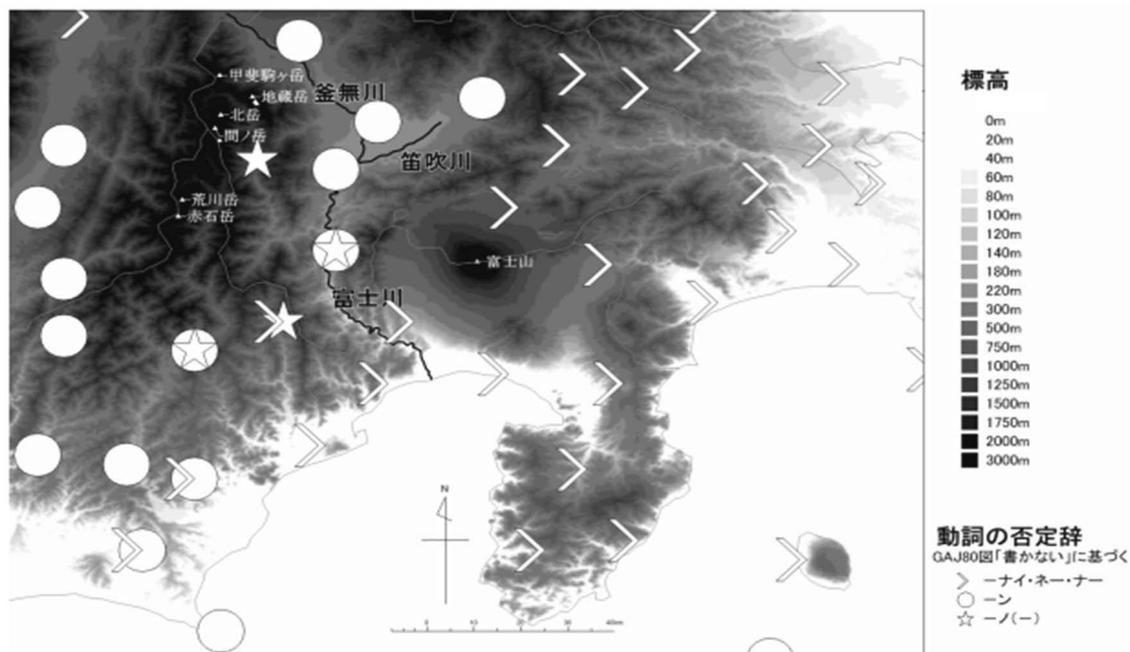


図5 富士川流域の動詞否定辞[16]

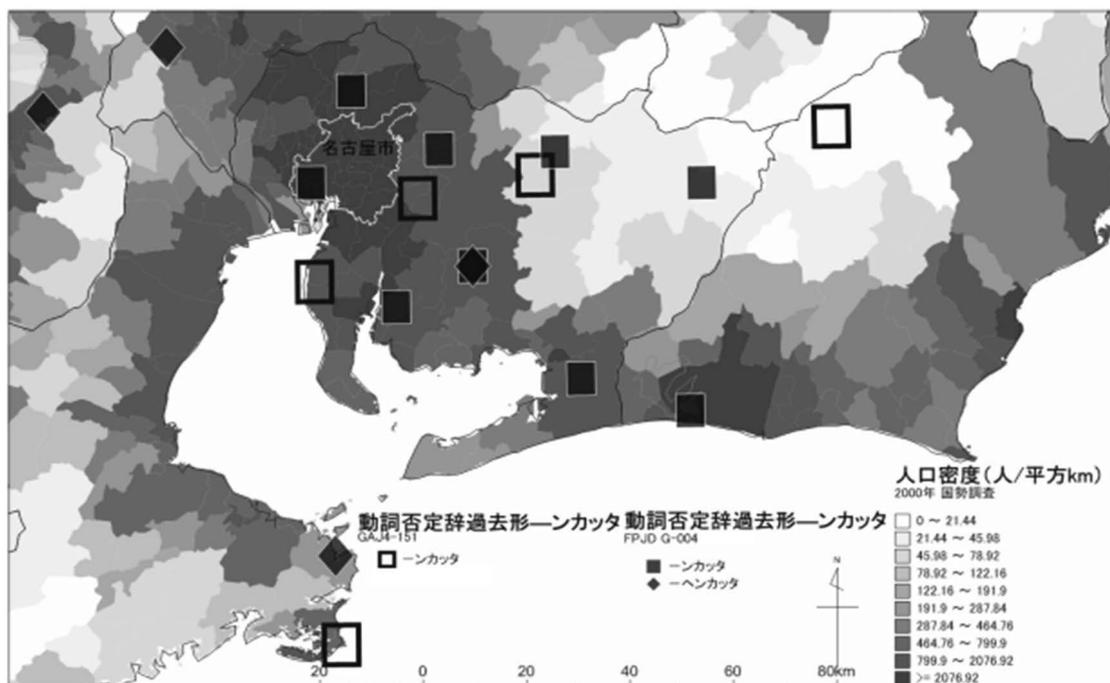


図6 東海地方のンカッタと人口密度[17]

## 9. むすび

方言分布・方言地図のデータベースについて、沿革と現状を述べた。方言学・言語地理学は、人文科学の中でも、保有・公開しているデータ量においては、かなり多い分野として、位置付けられるのではないかと思われる。

ただし、強調したいのは、そのような位置付けにあるわけではない。方言のデータには、場所を定義する空間情報が含まれている点である。空間情報をキーとすることで、分野間をリンクし、データの比較・対照・照合が可能になる。

その際に大きな力を発揮するツールは地図(そしてそれを実現するための GIS)である。ことば(方言)の情報と、例えば、歴史・地理・民俗などの情報の関係について、地図を用いて、また、地図の上で、客観的な裏付けに基づく分析・研究が展開できる。

方言学・言語地理学の中でも、方言の起源と形成過程の解明を目標とする方言形成論は、究極の課題であり、かつ言語に関する研究以外の分野との連携が必須である[18]。ここでは十分に触れることができなかったが、時間データも活用することで領域形成論[19]の再活性化も射程に収まるはずだ。ぜひ、諸分野からの意見や希望をお寄せいただきたい。

## 参考文献

- [1] 徳川宗賢. 架空教室「言語地理学演習」第一日. 言語, 1978, vol. 7, no. 9, p. 12-21.
- [2] 国語調査委員会. 音韻分布図. 日本書籍, 1905. 国語調査委員会. 口語法分布図. 国定教科書共同販売所, 1906.
- [3] 国立国語研究所編. 日本言語地図(全6巻). 大蔵省印刷局, 1966-1974.
- [4] 国立国語研究所編. 方言文法全国地図(全6巻). 大蔵省印刷局・財務省印刷局・国立印刷局, 1989-2006.
- [5] 大西拓一郎編. 新日本言語地図—分布図で見渡す方言の世界—. 朝倉書店, 2016.
- [6] 柴田武. 言語地理学の方法. 筑摩書房, 1969.
- [7] 柳田国男. 蝸牛考. 刀江書院, 1930.
- [8] 大西拓一郎. 方言地理学の研究動向. 方言の研究, 2016, no. 2, p. 83-97.
- [9] “『日本言語地図』データベース(LAJdb)”. <http://www.lajdb.org/>, (参照 2018-01-31).
- [10] 三石泰子. 言語地図集の目録. 言語生活, 1975, no. 284, p. 50-56.

- 三石泰子. 言語地図集の目録. 国語学, 1979, no. 119, p. 87-91.
- [11] “日本方言研究会”. <http://dialectology-jp.org/>, (参照 2018-01-31).
- [12] 大西拓一郎. 日本の方言地図と方言文法地図の動向. 日本語学, 1992, vol. 11, no. 6, p. 188-200.
- [13] 大西拓一郎. 主要方言地図目録. 馬瀬良雄監修. 方言地理学の課題. 明治書院, 2002, p. 446-465.
- [14] “方言研究の部屋”. <http://www2.ninjal.ac.jp/hogen/>, (参照 2018-01-31).
- [15] “方言の宇宙”. <http://www2.ninjal.ac.jp/takoni/>, (参照 2018-01-31).
- [16] 大西拓一郎. ことばの地理学—方言はなぜそこにあるのか—. 大修館書店, 2016.
- [17] 大西拓一郎. 言語変化と方言分布—方言分布形成の理論と経年比較に基づく検証—. 大西拓一郎編. 空間と時間の中の方言—ことばの変化は方言地図にどう現れるか—. 朝倉書店, 2017, p. 1-20.
- [18] 大西拓一郎. 方言形成論序説—言語地理学の再興—. 方言の研究, 2017, no. 3, p. 5-28.
- [19] 大林太良. 東と西 海と山—日本の文化領域. 小学館, 1990.

## 上智大学「日本の人名データベース」の構築について Constructing the Sophia University “Japanese Biographical Database”

ボルン 礼於

Leo Born

ハイデルベルグ大学 計算言語学部, Im Neuenheimer Feld 325, ハイデルベルグ, ドイツ  
Heidelberg University, Department of Computational Linguistics, Im Neuenheimer Feld 325,  
Heidelberg, Germany

**あらまし:** 上智大学の「日本の人名データベース」は、日本の歴代人物の伝記と、その個人を超える人間関係のネットワーク情報の提供を目的として作成された。本データベース構築は、頼春水(1746~1816)の研究から始まり、現在は人物約2,200人、人物間のイベント約7,000件を年月日・身分・出典などで検索し閲覧することができる。ハーバード大学のChina Biographical Databaseの構築をモデルにして、本データベースの改善の上、データベースをアクセスするウェブアプリケーションの開発をすすめている。将来的に人物のデータを蓄積するとともに、各人物のもつ姻戚・交友関係を詳細にモデル化することで、通史的検討に止まらない社会変遷の模様を追っていきたい。本報告では、本データベース構築の概要と、様々なオープンソースツールを用いたウェブアプリケーションの作成及びネットワーク可視化方法、今後の課題について述べる。

**Summary:** The Sophia University “Japanese Biographical Database” (JBDB) was constructed with the purpose of providing bibliographical information on Japanese historical figures and their personal, social, and political networks. Starting with research on Rai Shunsui (1746-1816), the database currently encompasses entries on ca. 2,200 individuals and ca. 7,000 events pertaining to these individuals and their interactions, allowing to search and visualize these entries by date, social status, sources, and other filters. Building upon the architecture of the Harvard University *China Biographical Database*, we are currently refining our own database, while also developing a new web application to access the database. With the addition of further biographical data, we aim to study patterns of societal changes that emerge out of familial and social relations among people. This paper focuses on the database design, the development of the web application with particular emphasis on the visualization component, and future directions for the project.

**キーワード:** データベース, ネットワーク研究, 集団履歴調査, 可視化

**Keywords:** database, network studies, prosopography, visualization

### 1. はじめに

現在、本研究グループは近世日本の人物についての伝記や、関係を取り扱うデータベースを構築し、あわせてデータベース上で動作される情報検索と可視化用のウェブアプリケーションを作成中である。本研究グループでは社会の構成要素として「集団」に着目し、その実態をとらえるため、集団履歴調査(プロソポグラフィ)やネットワーク分析(Social Network Analysis)の手法を検討している。歴史学における先行研究では、主に欧州史[1]や中国史[2]に焦点が置かれているが、本デー

タベースは、これまで集団履歴調査を含む分析の手法があまり適用されていない近世日本という研究対象にしている。これにより、近世日本の研究にネットワーク分析が導入される契機になるのではないかと考えている。

本プロジェクトの第一歩は、儒学者頼春水(広島藩儒)を取り巻くネットワークに着目することである。具体的には、ネットワーク分析や集団履歴調査の手法を用いた分析を行うため、頼春水の日記や書簡など、史料上で確認できる人名のデータベースを作成する。本プロジェクトは、個人の思想

トップ階層	テーブル数	件数
人物	5	2,209
文通	5	1,918
集会	4	4,982
アクティビティ	3	6,736
場所	8	5,699
文献	1	27

表 1. データ移動以前のデータ数

的評伝の形成のみならず、近世の人物の交流の様態をネットワーク分析の手法を用いて解明することで、伝記的研究からだけでは見えてこない同時代の学者交流および知の体系の蓄積像を明らかにしようとするものである。

そのため、過去 5 年にわたって人物数約 2,200 人、人物間のイベント数約 7,000 件がデータベースに入力された。このデータベースをもとに、ある人物が行動する時代的・場所的背景、またその人物が社会的にはたす役割を分析することが可能である。

本報告では、上智大学の「日本の人名データベース」の構築やウェブアプリケーション、およびネットワーク可視化ツールの概要について述べてみたい。

## 2. データベース構築

本プロジェクトは日本の歴代人物の伝記と、個人を超える人間関係のネットワークに関する情報を提供することを目的としているため、データの管理システムとして PostgreSQL[3]の関係データベースを用いた。データ入力は 27 件の参考文献に基づき手動で行っているが、デジタル化された文献を取得できる場合も考慮し、自動的な情報抽出方法も検討中である。

始めは、6 つのデータ内容で分類されたテーブル階層を中心に、合計 26 データベーステーブルを作成し、5 年間の期間でそこにデータが入力されてきた。その 6 つのトップ階層の統計を表 1 で示す。

しかし、データ記入の作業によって、特に件数が多いアクティビティという階層に問題があることが明らかになった。アクティビティというのは個人の家族、仕事、あるいは趣味に関わる情報をまとめて集めたテーブルである。そのためデータ属性や必然的なキーが大幅に異なっている。この問題を解消するために、データベース構築を改善し、データ移動を重点化することが決定された。

東亜史学圏でパイオニア的な役割を持つ、ハーバード大学の China Biographical Database (CBDB、



図 1. トップページ

「中国伝記データベース」)[4]をモデルにして、本データベース改善をすすめている。2016 年後期にハーバード大学から CBDB のデータベーステンプレートを頂き、それをもとに今までのデータを新しい構築にマッピングし、データ移動を行なっている。具体的には、データ移動以前の 26 テーブルのデータを、CBDB のテンプレートに基づいた 90 テーブルに移す作業である。

今まで表現の面で限界があった点も細かく区別をできるようになり、特に家族関係と仕事に関するテーブルが最も詳細である。その他にも、今までアクティビティで保存された住所変更や、身分変更も専用テーブルに記入できるようになる。今までモデル化されていなかった物事の購入・交換・獲得、又は執筆作品についての情報も、新データベース構築で適切にモデル化されている。

更に、「イベント」というものも、個人のイベントと数人が関係するイベントに適当にモデル化されている。それによって、各データカテゴリーに応じた属性を定義することができ、クエリと可視化に関する表現力が拡張される。

データ移動と同時期に、データベースにアクセスするウェブアプリケーションをモダン化する必要性に応じて、新ウェブアプリケーション開発を開始し、2017 年秋に「日本の人名データベース」(JBDB)として公開した[5]。以下でそのウェブアプリケーションについて述べてみたい。

## 3. ウェブアプリケーション

データベースはインターネット上で公開するものであるため、Node.js で動作する JavaScript のウェブアプリケーションを実装している。オープンソースツールを利用し、バックエンドは Express.js に基づく LoopBack[6] というフレームワーク、フロントエンドは AngularJS 1.X[7] で実行されている。



図 2. 人物のリスト



図 3. 人物の詳細ページ

LoopBackはオブジェクト関係マッピング (ORM) の導入と、REST APIの作成を容易にするため、データベース用とサーバー用のクエリデザインを統一させている。AngularJSではプラグインを取り込み、様々な機能を提供することができる。UI開発に主として利用しているプラグインはAngularJS制御用BootstrapのUI Bootstrap[8]である。多様な端末に対応させるため、レスポンシブデザインやWebフォントを可能にするフレームワークを導入することが重要である。

ウェブアプリケーションはトップページ (図1) にアクセスするとデータベースUIへのリンクと、スクロールダウンすると、研究グループの簡単な紹介が表示される。アプリケーションUIは2ヶ国語 (日本語と英語) であるため、スイッチ (トップページでは右上に配置) をクリックすることでUI言語を変更することができる。既定言語は日本語であるが、変更の場合、他の設定と一緒に言語の設定はクッキーに保存される。

データベースUIは既定では人物リストを表示し (図2<sup>(1)</sup>)、メニューバーでは人物、イベント (現在、文通と集会を含む)、出典のリストを選択できる。イベントと出典も、以下で述べる人物の表示と同様、基本情報が記載されているリスト表示と、詳細ページに分かれている。

人物リストで示されている情報は漢字とローマ字の書き方で名前、出身地 (藩・国・郡・村)、生まれの身分 (武士と非士の身分を合わせて18の選択肢)、生没年、データの更新日時、そしてログインユーザーのみが選択できるアクションである。生没年はUI言語の設定により、和暦又は西暦での表示である。生没年をマウスでホバーすると相互の紀年法が表示される。

データの閲覧にはログインすることは不要だが、データ編集にはユーザーアカウントが必要で

ある。現在、プロジェクトメンバーのみにアカウントがあり、そのメンバーには特別なアドミン権限が与えられている。なお、大学の授業でも活用していくために (5「おわりに」参照)、学生用の制限付きのアカウントも準備中である。

画面のヘッダーにはウェブアプリケーションのロゴと題名以外に、アプリケーション全画面に共通のナビゲーションメニュー (右側) を配し、ログインユーザー専用のメニュー (左側) をクリックでドロワーとして提供できるようにしている。

各データリストについては属性に対する簡易検索機能が用意され、リスト表示の対象を絞り込むことができる。例えば人物データは人名、または生まれの身分で検索することが可能である。検索結果を保存する簡易機能もあるが、現在.txtフォーマットに限られている。<sup>(2)</sup>

表示されたリストの項目をクリックすると閲覧ページに飛ぶ形態となっている (図3)。人物の場合は、表示される情報は三つの画面に分けてあり、既定では基本情報が表示される。データリストに記載されている情報以外に、人物の画像、他名のリスト (通称や諱)、データ入力に参考された文献のリスト、そしてコメントボックスが表示されている。基本情報の画面から関連するイベント (その人物が関わっているイベントのリスト)、そして履歴と家族関係に関わる情報が記載されている画面に飛ぶ仕組みであるが、後者のデータはまだ移動中であるため、非公開である。編集作業も図4に似た画面で行われ、多くの属性 (身分や出身地) はドロップダウンリストを利用し、入力する。

UIの目標はスムーズなUXを提供することであるため、多数のサードパーティライブラリーを加えたコードデザインを用意した。以上述べた通り、中心はBootstrapだが、それ以外に導入されているプラグインは表2で示す。

プラグイン名	機能
bootstrap-drawer [9]	メニューオーバーレイ用
angular-translate [10]	多言語 UI 用
angular-table [11]	ソート機能付きテーブル用
AngularJS Dropdown Multiselect [12]	ドロップダウン用
ng-file-upload [13]	ファイルアップロード用
Lightbox2 [14]	画像表示用
Angular File Saver [15]	検索結果保存用
Angular-xeditable [16]	インライン編集用
Animate.css [17]	一般 UI アニメーション用
AngularJS Slider [18]	スライダーUI 用

表 2. 利用中のサードパーティープラグイン

## 4. ネットワーク可視化

データベース構築の目的は、個人を超える情報の上で、集団ネットワークを明らかにすることである。そのため、様々なネットワーク的アプローチの調査手法を研究ツールとして提供し、インタラクティブなネットワークを可視化させる必要がある。

本ウェブアプリケーションのコードデザインの特徴は、基盤となるデータベースレイヤー上で行われる、モジュール化された多様コンポーネントの総合化にある。ネットワーク可視化は、モジュール化されたコンポーネントの一例として挙げられるが、地理情報システム (GIS) 等もそのようなコンポーネントとして考えられる。

本ウェブアプリケーションにおける、ネットワーク可視化ツールのコンポーネントは、オープンソースである vis.js [19] という可視化ライブラリーを用いて実装している。D3.js [20] 等の他の可視化プラグインよりも、vis.js は機能が少ないように見えるが、ネットワークとタイムラインの可視化を中心としているため、本プロジェクトには最適である。

また、ネットワーク可視化において、vis.js は様々なカスタマイズのオプションを提供しており、多様な物理的シミュレーションアルゴリズムを導入しており、特にダイナミックなネットワーク可視化用に便利である。データベース内のデータ量は今後拡大されるため、全ネットワーク共通のノード数とエッジ数は確定数ではない。



図 4. ネットワーク可視化ツールのフロントページ

### 4.1. 可視化ツールの使い方

ネットワーク可視化ツールのネットワーク画面の説明に具体例として、頼春水の息子である頼山陽 (1780~1832) のネットワーク可視化を取り上げてみたい (図 5)。

可視化ツールはフロントページ (図 4) にアクセスすると可視化設定画面が示される。その設定画面において、人物選択の手法とネットワークの表示に関する設定が行われる。

人物選択には現在以下の三つの手法がある。選択された手法によって、ネットワークは異なる人物に基づいて可視化される。

一) 「全員」を選択すると、データベース内の全人物が考慮される。<sup>(3)</sup> この設定ではデータベースの完全ネットワークが作成される。

二) 「名前で選択」を選択すると、人物検索窓と人物リストが示される。そのままリストから人物を選ぶこと、又は検索にて人物リストを絞り込むことで、考慮される人物が絞り込まれる。この設定では、選択された人物から始まり、その人物との関係性を持つ他の人物がデータベースからクエリされる。ネットワーク可視化はクエリ結果と元々選択された人物のセットの和集合に基づく。図 5 はこの設定で「頼山陽」を検索して選んだことに基づき作成された。

三) 「出典で選択」を選択すると、出典のドロップダウンが表示される。既定設定としては、選択された出典に伝記情報が書かれてある人物から始まり、その人物との関係性を有する他の人物がデータベースからクエリされる。ネットワーク可視化はクエリ結果と選択された出典に現れる人物のセットの和集合に基づく。

現在、クエリで人物の間関係性を定めるには、イベント (文通と集会) のみが考慮される。しかし、今後はネットワークを家族関係や仕事関係の情報にも基づくようにする予定である。

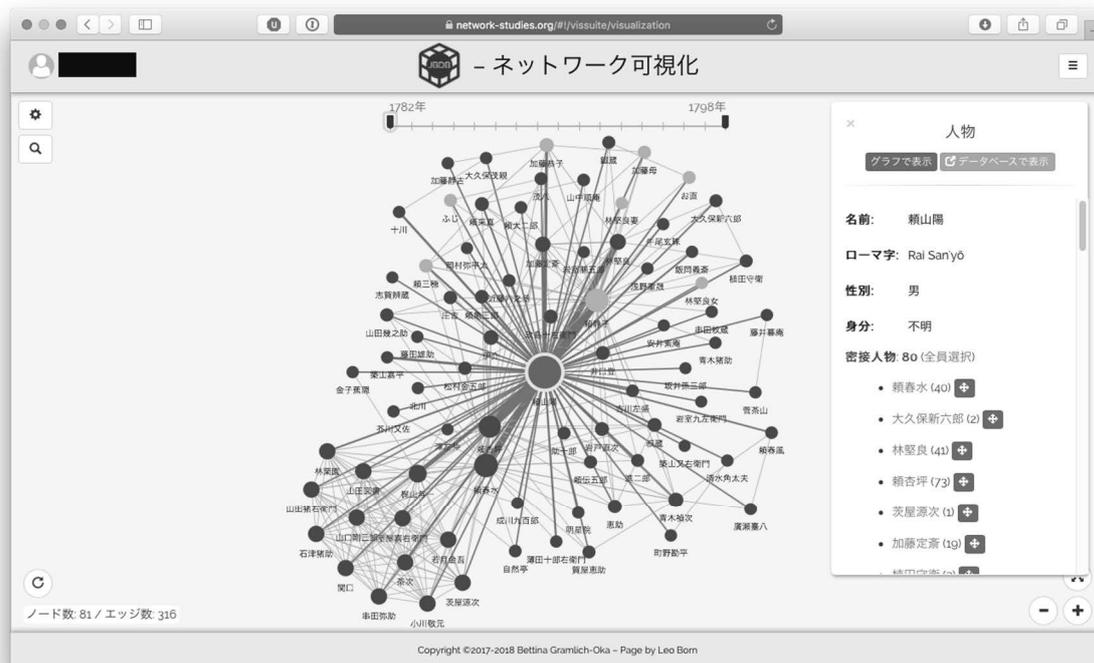


図 5. 賴山陽のネットワーク可視化

人物の選択手法が決定された後、ネットワークでノードの附属グループをどの属性に基づいて強調表示するか選択できる。附属グループを示す属性は現在、性別と生まれの身分のみだが、データベース内のテーブル属性に任意に基づくため、他の属性も今後追加することができる。図 5 は性別のグループ附属にて作成された。(4)

可視化のレンダリングが終了した状態は、図 5 で示している。画面の左下にはノード数とエッジ数を示すオーバーレイがある。ネットワーク自体はインタラクティブで、自由にノードを動かす、ズームレベルや表示フレームも自由に変更することが可能である。

ひとつのノードを選択すると、繋がっているノードとそのエッジが強調される上に、選択されたノードの詳細情報が別のオーバーレイで表示される。オーバーレイは人物の基本情報と、どの他の人物と繋がっているかをリストで示し、その関係人物ををクリックで全員選択して、グループとして動かせる。さらに、密接人物のリストからは、他の人物との共通イベントを Bootstrap モーダルで表示し (図 6)、画面フレームを他の人物の方に動かすことができる。人物オーバーレイやモーダルリストから新しいタブで、直接データベースでのエントリーに飛ぶ形態となっている。ネットワークを変更状態から初期状態に戻すには、画面の左下

のリセットボタンをクリックする。最後に、画面の左上の検索メニューによって、ネットワークで表示されている人物を検索することができる。

以上で述べた基本設定以外の複数の追加設定 (図 7) について、次で説明していきたい。

#### 4.2. 共通追加設定

この共通追加設定は、人物の選択手法に関わらず選択できるものである。図 5 は追加設定の「接合強度を可視化」、「人物の重要性を可視化」、「タイム・スライダーを追加」で作成された。

最初の設定、「接合強度を可視化」では、ネットワークで二人の人物の間の接合強度が可視化される。接合強度とは、ある二人の人物の間の接合数 (現在、共通のイベント数) にて定められる。接合数が多い人物ペアの間のエッジは接合数が少ない人物ペアの間のエッジより太くレンダリングされる。

次の設定、「人物の重要性を可視化」では、ノードの大きさが人物の重要性 (接合度) によって可視化される。重要性とは、ある人物のネットワークでの次数 (エッジ数) で定められる。接合数が多い人物は接合数が少ない人物より大きく示される。この設定が選択されない場合でも、可視化画面の設定メニュー (歯車シンボルをクリックしてアクセス) でスライダーを使い、ネットワークでの人物



図 6. ネットワークで選択した人物の詳細モーダル

を最低字数で強調表示することができる。この設定を選択せずに、次のタイム・スライダーの設定を選択すると、次数スライダーとタイム・スライダーの使用は選言的となる。

第3の共通追加設定、「タイム・スライダーを追加」では、ネットワークでのエッジ（現在、イベント）を年でフィルターできるスライダーが可視化画面に追加される。このタイム・スライダーを使うことで、ネットワークの構造を年ごとに分析することができる（図8）。

最後の共通追加設定、「グラフで未接合のノードを表示」では、エッジのないノードも表示される。この設定により、データベース内で他の人物と関係ない人物も考慮させることができる。人物全員とこの設定を選択すると、データベース内の全人物がネットワークでレンダリングされる。

共通設定とは言え、追加設定の使用やコンビネーションには制限がある。選択手法ごとに述べてみたい。「全員」の選択手法では、タイム・スライダーが低速であるため、利用不可である。ノード数、又データベース内のイベント年の範囲において、計算が複雑になるためである。

「名前で選択」の手法では、未接合のノードを表示することが不可能である。この設定のクエリ構築では、選択された人物と必ず関係を持つ他の人物のみが考慮されるため、未接合のノードは発生しない。

最後に、「出典で選択」の手法では、タイム・スライダーと未接合ノードの追加設定は選言的である。タイム・スライダーを表示することで、ネットワークをイベントの日時で閲覧するため、イベントなしの未接合ノードは無用となる。また、未接合ノードを示す設定では、日時でフィルターすることが困難である。

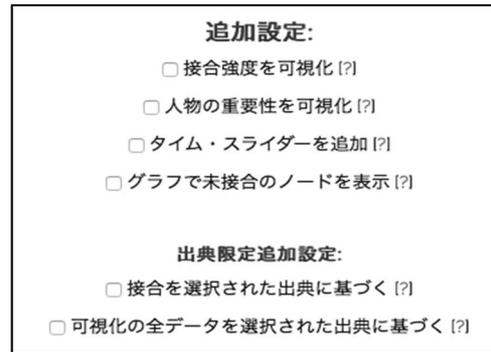


図 7. ネットワーク可視化ツールの追加設定

#### 4.3. 出典限定追加設定

以上で述べた共通の追加設定以外に、「出典で選択」の手法に限られる、二つの追加設定がある。この設定では、出典の範囲を確定することによって、ネットワークを段階で出典中心的に作成することになるので、この追加設定は選言的である。

一段目は、出典限定追加設定が選択されていない、デフォルトの状態である。この段階では、これまで述べた通り、選択された出典に伝記情報が記される人物から始まり、その人物との関係性を有する人物がクエリされ、クエリ結果と選択された出典に現れる人物のセットの和集合に基づき、ネットワークがレンダリングされる。

次の段階は、「接合を選択された出典に基づく」の第1の選択肢である。この設定では、ネットワークのエッジ（接合）に基づくデータが選択された出典に限られる。選択された出典に基づくイベントに参加する人物のセットと、既定設定と同様に、選択された出典に伝記情報が載る人物のセットの和集合に基づき、ネットワークが表示される。このような仕組みでは、伝記情報が記されてある人物だけではなく、イベントのコンテキストのみで選択された出典で現れる人物も考慮される。

最後の段階は、「可視化の全データを選択された出典に基づく」の第2の選択肢で、出典において一番狭い範囲でネットワークを作成する設定である。この設定では、ネットワーク全体に基づくデータが、選択された出典に限られる。選択された出典に伝記情報が記されてある人物のみが考慮され、更にある二人の人物の間の接合も選択された出典に限られて定められる。この設定では、選択された出典のみのデータがネットワークのインプットとなるため、出典自体がもたらすネットワークが可視化されることになる。

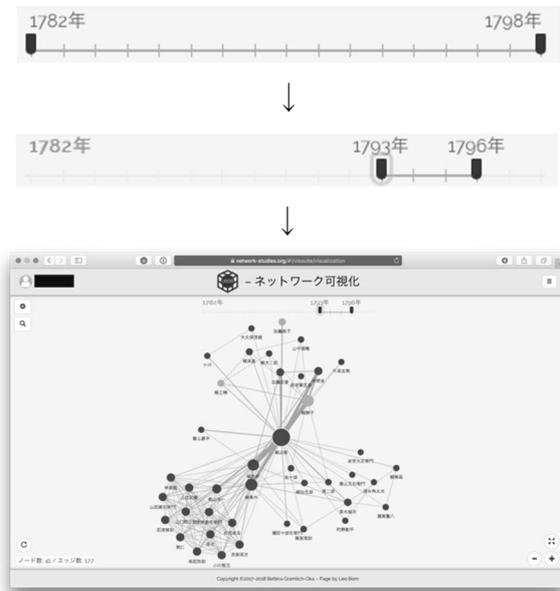


図 8. 図 5 でタイム・スライダーを使用した結果

## 5. おわりに

本プロジェクトの一部をなす「近世日本の知的ネットワークの研究」では、頼春水のグループを含め、様々な集団に焦点を当てる。そして、ネットワークにかかわった人々の書きものを幅広く取り上げ、総合的な分析を加えていく。人物が藩政や国政にかかわっている場合、各藩内また藩の枠を超えて広がる思想的ネットワークを研究することは、近世日本で政治改革がどのように論じられ、実施されたかを理解する重要な手掛かりとなるだろう。ネットワーク的アプローチは、近世日本の政治が幕府、諸藩、またそれ以下のレベルで、具体的にどう動いていたのかを理解する上でも有用な見方を提供してくれる。

データベースの構築の上、ウェブアプリケーション及びネットワーク可視化ツールの運用によって、この目標に近付くことができる。ウェブアプリケーションのデザインの利点としてモジュール化されたコンポーネントの導入が挙げられる。今後、ネットワーク可視化用のコンポーネント以外にも、空間的なデータの分析に用いられる地理情報システム (GIS) をコンポーネントとして取り入れる見込みである。

さらに、本研究プロジェクトでは、長期的には同様の研究手法を用いて、近世日本以外にも異なる時代や地域を対象とした研究を進めていく予定である。このような時空間情報及び文化横断的アプローチに基づくツールの提供は研究だけでなく、教育にも役立つだろう。

## 注釈

- (1) 図表ではログインされた状態が表示されるため、画像は編集されている。
- (2) CSV と XML フォーマットでも保存できることを予定している。
- (3) この点は可視化の追加設定にて制限される場合もある (項 4.2 参照)。計算の複雑性とパフォーマンスのため、既定では未接合のノード (データベース内で他の人物と関係ない人物) が表示されないが、追加設定にて未接合のノードも考慮させることができる。後者の方を選択する場合のみにデータベース内の全人物がネットワークで考慮される。
- (4) グループ数は属性にて異なり、確定ではないため、強調用の色の選択はランダム化されている。

## 参考文献

- [1] Padgett, J. F. and Ansell, C. K.. *Robust Action and the Rise of the Medici, 1400-1434*. The American Journal of Sociology, 1993, vol. 98, no. 6, p. 1259-1319.
- [2] Berman, M. L.. *Boundaries or Networks in Historical GIS: Concepts of Measuring Space and Administrative Geography in Chinese History*. Historical Geography, 2005, vol. 33, p. 118-133.
- [3] PostgreSQL. <https://www.postgresql.org/>, (参照 2018-01-30).
- [4] Harvard University, Academia Sinica, and Peking University. *China Biographical Database*. <https://projects.iq.harvard.edu/cbdb/>, (参照 2018-01-30).
- [5] Sophia University. *Japanese Biographical Database*. <https://network-studies.org/>, (参照 2018-01-30).
- [6] LoopBack. <https://loopback.io/>, (参照 2018-01-30).
- [7] AngularJS. <https://angularjs.org/>, (参照 2018-01-30).
- [8] UI Bootstrap. <https://angular-ui.github.io/bootstrap/>, (参照 2018-01-30).
- [9] bootstrap-drawer. <https://github.com/clineamb/bootstrap-drawer/>, (参照 2018-01-30).
- [10] angular-translate. <https://angular-translate.github.io/>, (参照 2018-01-30).
- [11] angular-table. <https://github.com/samu/angular-table/>, (参照 2018-01-30).

- [12] AngularJS Dropdown Multiselect. <http://dotansimha.github.io/angularjs-dropdown-multiselect/docs/#/main>, (参照 2018-01-30).
- [13] ng-file-upload. <https://github.com/danialfarid/ng-file-upload>, (参照 2018-01-30).
- [14] Lightbox2. <http://lokeshdhakar.com/projects/lightbox2>, (参照 2018-01-30).
- [15] Angular File Saver. <http://alferov.github.io/angular-file-saver/>, (参照 2018-01-30).
- [16] Angular-xeditable. <https://vitalets.github.io/angular-xeditable/#>, (参照 2018-01-30).
- [17] Animate.css. <https://daneden.github.io/animate.css/>, (参照 2018-01-30).
- [18] AngularJS Slider. <http://angular-slider.github.io/angularjs-slider/>, (参照 2018-01-30).
- [19] Almende B.V. *vis.js*. <http://visjs.org>, (参照 2018-01-30).
- [20] D3.js. <https://d3js.org>, (参照 2018-01-30).

第23回公開シンポジウム「人文科学とデータベース」発表論文集

2018年3月3日発行

編集・発行

第23回公開シンポジウム「人文科学とデータベース」実行委員会

事務局 〒599-8531 大阪府堺市中区学園町1-1

大阪府立大学人間社会システム科学研究科

小島篤博研究室内

ISSN 2188-1529