

第25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 25th Symposium  
on Humanities and Database

2020年2月29日(土)

図書館流通センター

本社ビル 地下ホール

【主催】 第25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、TRC-ADEAC 株式会社

【後援】 アート・ドキュメンテーション学会、情報知識学会、

情報処理学会人文科学とコンピュータ研究会

第25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 25th Symposium  
on Humanities and Database

2020年2月29日(土)  
図書館流通センター  
本社ビル 地下ホール

- 【主催】 第25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、TRC-ADEAC 株式会社  
【後援】 アート・ドキュメンテーション学会、情報知識学会、  
情報処理学会人文科学とコンピュータ研究会

【協賛】

SABIA 



株式会社 インフォマーヂュ

SIP

文化事業・情報処理・サイン・複写  
Shinko Industrial Photo Corporation  
四国工業写真株式会社

株式会社  
TRC図書館流通センター

ADEAC®



## 目 次

### ●一般講演 1

日本語研究分野における研究動向の 70 年

－ 国立国語研究所「日本語研究・日本語教育文献データベース」による－

八木下孝雄、高田智和（国立国語研究所）……………1

国立国語研究所研究資料室収蔵資料と目録検索

高田智和、関川雅彦、山口亮、石本祐一（国立国語研究所）……………7

### ●一般講演 2

人文科学系データベースにおけるユーザビリティの検証

－ Google Chrome 拡張機能を用いた取り組み－

青柳明佳（国文学研究資料館）、篠原泰彦（朔工房）……………9

日本語学習ディクテーション・データベースの開発

山元啓史（東京工業大学）、ホドシチェク ボル（大阪大学）……………17

### ●基調講演

歴史知識における「時」の表現と構造

赤石美奈（法政大学・情報科学部 教授）……………25

### ●企業システム紹介

デジタルアーカイブシステム ADEAC の紹介

田山健二（TRC-ADEAC 株）……………33

「伊能図」の測量技術

太田武和（測量士補、元伊能忠敬記念館説明員）……………34

●一般講演 3

- 「伊能大図」に対する GIS データの活用  
菅原真悟 (TRC-ADEAC 株) .....37
- 森鷗外の小説を対象とした文体の継時的な変化についての計量的な検討  
土山玄 (お茶の水女子大学) .....45
- ADEAC の画像データを利用したくずし字認識 AI の開発と組み込みシステムへの実装  
早坂太一、竹内正広、大野互 (豊田工業高等専門学校)、加藤弓枝 (鶴見大学)、  
山本和明 (国文学研究資料館)、石間衛、石川徹也 (TRC-ADEAC 株) .....53

●一般講演 4

- 無形民俗文化財のデータベース化のための情報の体系的記述に関する検討～浜松市における～  
杉山岳弘 (静岡大学) .....59
- 古天気、作況、死亡データベース構築の構想－ 18・19 世紀の東北地方における死亡危機の要因解  
明に向けて－  
川口洋 (帝塚山大学) .....67
- 精緻な表記情報を有する「延喜式祝詞」コーパスの構築  
間淵洋子 (国立国語研究所) .....75

# 一般講演 1

座長:田窪直規(近畿大学)



日本語研究分野における研究動向の 70 年  
—国立国語研究所「日本語研究・日本語教育文献データベース」による—  
**Research Trends on the Japanese Linguistics for 70 Years:  
From the Bibliographic Database of Research on Japanese Language  
and Japanese Language Education**

八木下 孝雄, 高田智和

Takao YAGISHITA, Tomokazu TAKADA

国立国語研究所, 東京都立川市緑町 10-2

National Institute for Japanese Language and Linguistics,

10-2 Midori-cho, Tachikawa City, Tokyo

**概要:** 2011 年からオンライン・データベースとして国立国語研究所で公開している「日本語研究・日本語教育文献データベース」(<https://bibdb.ninjal.ac.jp/bunken/>)は、日本語研究及び日本語教育分野の論文・図書の書誌情報目録データベースである。『国語年鑑』(1954～2009 年, 国立国語研究所), 『日本語教育年鑑』(2000～2008 年, 国立国語研究所)の文献目録を前身としており, 1950 年以降およそ 70 年間の研究成果を通覧できる。本データベースを用いて, 日本語研究分野(文法, 語彙, 文体などの研究分野)の研究動向を見ると, 1950 年以降 2006 年ごろまでデータ件数(登録されている論文・図書の件数)は増加傾向であるが, その後件数が減少傾向に転じていることや, 分野毎に増減の傾向に差があることなどがわかった。本研究では, 「日本語研究・日本語教育文献データベース」を用いて以上のような日本語研究分野のおよそ 70 年にわたる研究成果を概観していく。

**Abstract:** *The Bibliographic Database of Research on Japanese Language and Japanese Language Education* (<https://bibdb.ninjal.ac.jp/bunken/>, hereafter *the Bibliographic Database*) is a database of material related to the Japanese language and Japanese language education, which is published by the National Institute for Japanese Language and Linguistics. The database catalogues books and articles on these topics that have been published over the nearly 70 years from 1950 to the present. In this paper, we aim to find research trends in Japanese linguistics from the nearly 70 years of records from *the Bibliographic Database*. On analyzing research trends of Japanese linguistics, we find the results as stated below. (1) Data counts from 1950 to 2006 are on the increase, but also those from after 2006 are on the decrease. (2) In each research field our analysis did not show the same trends.

**キーワード:** 日本語研究, 「日本語研究・日本語教育文献データベース」, 研究動向

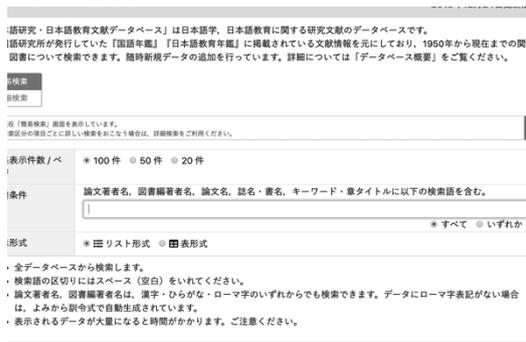
**Keywords:** Japanese Linguistics, *the Bibliographic Database of Research on Japanese Language and Japanese Language Education*, Research Trends

## 1. はじめに

国立国語研究所の「日本語研究・日本語教育文献データベース」(図 1, <https://bibdb.ninjal.ac.jp/bunken/>)は、国内外のユーザーに、日本語研究分野と日本語教育分野を中心とした学術文献の情報を提供している。2011 年に、『国語年鑑』と『日本語教育年鑑』の

論文データを統合して以来, 定期的なデータ追加と, さまざまな機能追加を行ってきた。

後述するが、「日本語研究・日本語教育文献データベース」は過去約70年の日本語研究に関する学術文献の情報が検索可能になっており、データを用いて研究動向を分析することが可能である。本発表では、「日本語研究・日本語教育文献データベース」のデータを分析することで、日本語研究がどのように変遷してきたのかを概観する。



【図1 「日本語研究・日本語教育文献データベース」簡易検索画面】

## 2. 「日本語研究・日本語教育文献データベース」について

「日本語研究・日本語教育文献データベース」(以下「文献DB」)は、(1)雑誌論文データ(学術雑誌・大学紀要等の掲載論文)、(2)論文集データ(図書としての論文集等の掲載論文)、(3)図書データの3種からなる複合データベースとなっている。総レコード数は2019年12月現在で、約27万件である。以下にその内訳を示す。

- |             |       |             |
|-------------|-------|-------------|
| (1) 雑誌論文データ | ..... | 約 192,000 件 |
| (2) 論文集データ  | ..... | 約 30,000 件  |
| (3) 図書データ   | ..... | 約 48,000 件  |
| 合 計         |       | 270,000 件   |

### 2.1 収録データの採録範囲について

文献DBは、前述したように、『国語年鑑』と『日本語教育年鑑』の文献情報を統合し、引き継ぐ形でWeb上での検索が可能なデータベースとして公開されてきた。

#### 2.1.1 『国語年鑑』

『国語年鑑』は、1954年から2009年まで国立国語研究所で刊行されていたもので、研究文献情報を中

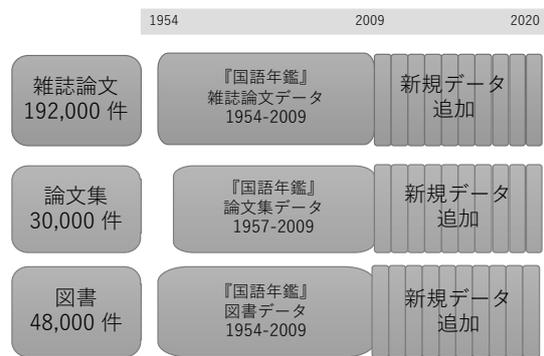
心に学界の情報が掲載されている。1954年(昭和29年)に刊行が始まり、冊子としての刊行は2008年で終了、2009年に電子版がPDFで刊行され、刊行が終了している。昭和29年版から2009年版まで全部で56冊刊行されている。

#### 2.1.2 『日本語教育年鑑』

『日本語教育年鑑』は、日本語教育に関する文献情報や学界の情報が掲載されており、2000年に刊行が開始され、2008年に刊行が終了している。なお、文献DBに採録されている『日本語教育年鑑』関係の文献情報は、その前身となる1980年から1998年まで刊行された『日本語教育学会誌・機関誌掲載論文等文献一覧』(国立国語研究所日本語教育センター・第二研究室)のデータも含まれており、1960年代の文献情報から収録されている。

#### 2.1.3 両年鑑刊行終了後のデータ

上記、『国語年鑑』および、『日本語教育年鑑』の刊行終了後の文献情報のデータ採録は、主に国立国語研究所研究図書室に受け入れた雑誌・図書から行っている。



【図2 「日本語研究・日本語教育文献データベース」データの構成<sup>1)</sup>】

### 2.2 データ項目について

文献DBのデータ項目を以下に示す。

1. DB
2. 文献ID
3. 研究図書室請求記号
4. 論文著者名
5. 論文著者名別表記
6. 論文名

<sup>1)</sup> 論文集論文データは、『国語年鑑』での掲載が昭和32年(1957年)刊行のものからである。

7. 論文名別表記
8. 図書編著者名
9. 図書編著者名別表記
10. 誌名・書名
11. 誌名・書名別表記
12. 巻号
13. ページ
14. 総ページ数
15. 発行
16. 発行年月
17. 発行国
18. キーワード
19. 章タイトル・目次
20. 分野
21. 論文著者名ローマ字(自動生成)
22. 図書編著者名ローマ字(自動生成)
23. 更新
24. (リンク)

#### 2.2.1 「分野」について

データはそれぞれ文献をもとに入力されているが、前節に示した「20. 分野」の項目は日本語研究の専門家が文献の内容を判断し登録している。「分野」は以下の通りである。

1. 日本語学一般
2. 日本語史
3. 音声・音韻
4. 文字・表記
5. 語彙・用語
6. 文法
7. 文章・文体
8. 方言
9. 日本語情報処理
10. コミュニケーション
11. マスコミュニケーション
12. 国語問題・言語問題
13. 国語教育
14. 日本語教育
15. 言語学
16. 資料
17. 書評
18. 辞書

なお、「分野」については、データ1件に対し、最大3つまでの分野が付与されている。これは、分野が複数にわたる研究もあるためである。

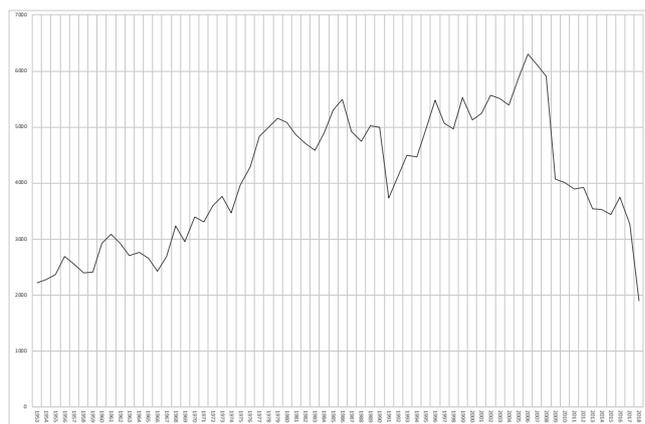
### 3. 研究方法

文献 DB のデータを用いて、日本語研究における研究動向を概観する。まず、全体的なデータ件数の推移を確認する。その後、データ項目から「分野」を中心に分析し、研究内容も含めた研究動向を見ていく。2.2.1 で述べたように、「分野」は日本語研究の専門家による文献内容の判断を経て付与されており、これを見ることで、日本語研究における研究の動向を見ることができると考えられる。なお、データ分析にあたっては、1953年～2018年発行の文献のデータを取り扱う。<sup>2</sup>

### 4. 調査・分析結果

#### 4.1 データ件数の経年変化

まず、データ件数の経年変化を確認する。



【図3 データ件数の経年変化】

【図3】を見ると、2006年までは件数が増加傾向であるが、その後減少傾向に転じ、最近10年間は年間3500件程度になっていることがわかる。近年の日本語研究の規模が縮小していることがうかがえる。

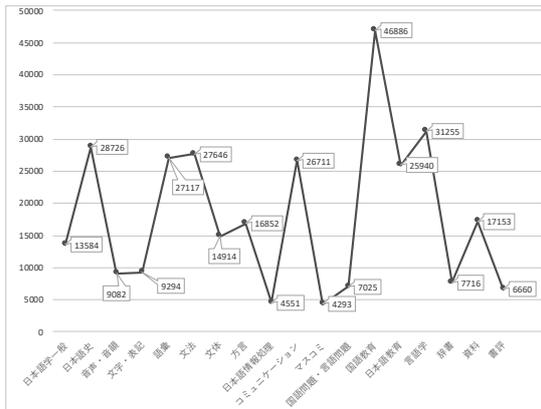
#### 4.2 分野ごとの分析

前述のように、分野は、データ1件に対して最大3つまで付与される。そのため、データ毎に分野を分析することが難しい。ここでは、分野1つに対して1データと考え、分析していく。

<sup>2</sup> 『国語年鑑』は刊行年の前年の文献情報の掲載が中心となる。そのため、初刊は1954年刊行であるが、掲載されているのは、1953年に発行された文献情報が中心となる。2019年刊行の文献についてのデータは現在入力作業中であり、

比較できるだけのデータ数が揃っていないため、分析対象から外した。

#### 4. 2. 1 分野ごとの総数



【図4 分野毎の総数】

【図4】は、分野ごとに件数を集計したものである。まず、件数が最も多いのは「国語教育」分野の研究であることがわかる。「言語学」、「日本語史」、「文法」、「語彙」、「コミュニケーション」、「日本語教育」の分野が25,000件以上で続く。15,000件前後のものとしては、「資料」、「方言」、「文体」、「日本語学一般」があり、その他の分野は10,000件以下となっている。

#### 4. 2. 2 分野毎の経年変化

分野毎の経年変化を確認する。それぞれグラフを以下に示す。



【図5 分野「日本語学一般」の経年変化】



【図6 分野「日本語史」の経年変化】



【図7 分野「音聲・音韻」の経年変化】



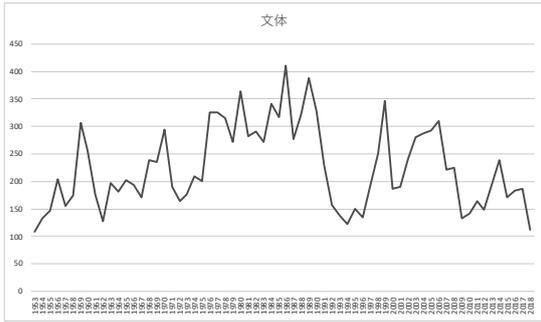
【図8 分野「文字・表記」の経年変化】



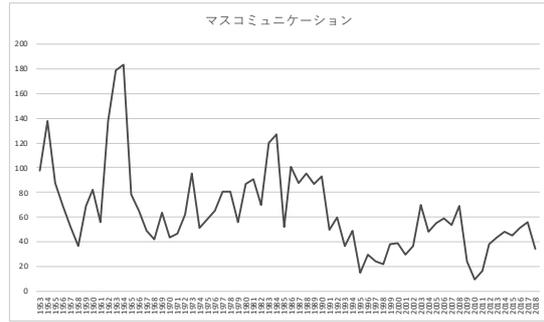
【図9 分野「語彙」の経年変化】



【図10 分野「文法」の経年変化】



【図 11 分野「文体」の経年変化】



【図 15 分野「マスコミュニケーション」の経年変化】



【図 12 分野「方言」の経年変化】



【図 16 分野「国語問題・言語問題」の経年変化】



【図 13 分野「日本語情報処理」の経年変化】



【図 17 分野「国語教育」の経年変化】



【図 14 分野「コミュニケーション」の経年変化】



【図 18 分野「日本語教育」の経年変化】



【図 19 分野「言語学」の経年変化】



【図 20 分野「辞書」の経年変化】



【図 21 分野「資料」の経年変化】



【図 22 分野「書評」の経年変化】

上記の図をそれぞれ見ていくと、全体のデータ件数の増減の傾向と一致しているのは、「文体」「方言」「辞書」であった。また、全体よりも件数の増加が早い時期に起こり、一旦減少するものの、近年再度増加傾向にあるものが、「日本語史」「音声・音韻」「語彙」「言語学」の分野である。上昇傾向のみのものが、「文字・表記」「文法」「日本語情報処理」「コミュニケーション」「日本語教育」であった。下降傾向のみのものは、「マ

スコミュニケーション」「国語問題・言語問題」「国語教育」であった。また傾向が不明なものとして、「資料」「書評」があった。このように、全体的な傾向に一致していないものが多く、分野毎に、研究が盛んな時期、沈滞している時期などが見て取れる。

#### 4.3 考察

【図4】から、過去約70年間の日本語研究で最も研究されている分野は「国語教育」に関するもののである。しかし、これはデータの前身となっている『国語年鑑』の文献情報の収集方針とも関連していると考えられる。刊行初期の『国語年鑑』は国語教育の文献情報を多数載せていた。[1]【図17】の「国語教育」分野の経年変化を確認すると、登録データについて一貫して減少傾向にあることが分かる。また、2番目に多く研究されていると考えられる「言語学」分野については、日本語と外国語の対照研究なども多く含まれるため、日本語研究の関連研究と捉えられるものもデータとして登録されている。

以上から、日本語研究そのものの研究としては「日本語史」がもっとも多く、「文法」、「語彙」、「コミュニケーション」と続いている。これらの4つの分野について、経年変化を確認すると、「日本語史」「語彙」は、増加傾向の後に減少し、近年再度増加しているのに対し、「文法」「コミュニケーション」は全体的に上昇傾向である。件数が同程度でも、研究の活況、沈滞により「分野」の経年変化に差が見られることがわかる。

#### 5. おわりに

文献DBのデータを用いて、日本語研究の研究動向を見てきた。以下に、まとめとして見られた傾向を示す。

- 1) 日本語研究は2006年を境に研究の件数が上昇傾向から下降傾向へ転じている。
- 2) 分野毎の件数では、「国語教育」、「言語学」の件数が多くなっている。その他に、「日本語史」、「文法」、「語彙」、「コミュニケーション」、「日本語教育」の分野が多く研究されている。
- 3) 分野毎にデータ件数の経年変化を見ると、全体のデータ件数の傾向と合致している分野の方が少なく、上昇傾向が続いているもの、減少傾向がつついているものなど、それぞれ分野により活況、沈滞、と傾向が分かれている。

#### 参考文献

- [1] 八木下孝雄(2018)『『国語年鑑』の分野区分にみる研究動向の変遷』『日本語学会2019年度春季大会予稿集』日本語学会, pp.159-162

# 国立国語研究所研究資料室収蔵資料と目録検索

高田智和、関川雅彦、山口亮、石本祐一

国立国語研究所，東京都立川市緑町 10-2

**概要：** 国立国語研究所は、1948 年の設立以来、日本語に関する様々な調査研究を行ってきた。研究成果はその都度、報告書や論文として公刊してきたが、研究成果の前段階の中間生成物に相当する情報カードや集計表、さらに、研究の一次資料に当たる調査票、録音、語彙調査の雑誌原本、調査研究運営の記録である調査計画書や会議録も現存し、それらは国語研研究資料室に保存して集中管理を行っている。概ね調査研究単位の資料群に整理し、資料群の概要記述(239 件)、紙資料の保存箱の目録(4,215 件)、音声映像資料のデジタル化ファイル目録(24,258 件)を作成し(国立国語研究所研究資料室収蔵資料、<http://rnr.ninjal.ac.jp/>)、全文検索サービスを提供している。音声・映像を含めて、収蔵資料は研究・教育目的であれば、来館利用が可能である。目録検索サービスは、資料の保存管理、過去の調査結果の検証可能性の確保、新たな研究課題での再利用を目的としたものである。



## **一般講演 2**

**座長:西村陽子(東洋大学)**



# 人文科学系データベースにおけるユーザビリティの検証 GoogleChrome 拡張機能を用いた取り組み Verification of usability in humanities database

青柳 明佳

Sayaka Aoyagi

国文学研究資料館 論文目録データベース室、東京都立川市緑町 10-3

National Institute of Japanese Literature, 10-3 Midori-cho, Tachikawa city, TOKYO 190-0014, Japan

篠原 泰彦

Yasuhiko Shinohara

朔工房

Sakukoubou

**概要:** Web 上で検索できるデータベースは、様々な調査・研究を行い、論文を執筆するために、もはや必須のツールと言っても過言ではない。だが、データベースを構築・運用するにあたり、作成者の意図と利用者の目的や使用方法が必ずしも一致するとは限らない。また、一度稼働し始めたデータベースは、それを止めて作り直す、あるいは修正を行うことが難しい。そこで、本研究では、2019年10月時点で64.92%のシェアを誇るブラウザ「Google Chrome」の拡張機能を使用し、既存の人文科学系論文データベースである CiNii Articles、NDL ONLINE、国文学論文目録データベース、日本語研究・日本語教育文献データベースを例に、既存のデータベースの運用を止めることなく、表示方法や操作方法を変えながら「データベースにおけるユーザビリティ」を検証していく。また、この方法を通して得られた「どのような点に留意してデータベースの構築・運営を行うべきか」の知見を提示したい。

**Abstract:** A database that can be searched on the Web is no longer an indispensable tool for conducting various surveys and researches and writing dissertations. However, in building and operating a database, the intention of the creator does not always match the purpose and usage of the user. Also, once the database has started running, it is difficult to stop and recreate or modify it. Therefore, in this study, we used the extension function of the browser "Google Chrome" which boasts a share of 64.92% as of October 2019, and used the existing humanities-related dissertation databases CiNii Articles, NDL ONLINE, Japanese literature catalog database, Japan Using the example of a language research and Japanese language education literature database, we will continue to verify "usability in a database" by changing the display method and operation method without stopping the operation of the existing database. In addition, I would like to present the knowledge obtained from this method on "what points should be considered when constructing and operating a database"

**キーワード:** 人文科学, データベース, ユーザビリティ, 拡張機能

**Keywords:** humanities, database, Usability, Extension

## 1. はじめに

論文を作成するには、先行論文を収集し、それらの資料をふまえた考察、検討が不可欠である。資料収集

の方法としては、手当たり次第に探してみる、手持ちの文献資料から参考文献をたどる、対象分野のデータベースを使い資料を検索する、といった方法がある。

手当たり次第に探す方法は、すぐ確認できる反面、網羅的に探すことが難しい。また、手持ちの文献資料から参考文献をたどる方法は、求めている文献が見つかる可能性も高いが、その資料より古いものしかないというデメリットがある。データベースを使って検索する方法は、OPAC等を用いて検索を行い、必要があればそこから機関ポータルにアクセスして資料を閲覧したり、原本の複写を申請したりできるため、効率的である。そして、研究機関や個人の手によって様々なデータベースが構築・運用されるようになった。

人文科学の分野に限っても、さまざまな機関で独自のデータベースが作られている。国内の人文系論文データベースとしては、文学研究資料館の「国文学論文目録データベース」、「アーカイブズ学文献データベース」、国立国語研究所の「日本語研究・日本語教育文献データベース」、「Web データに基づく用例データベース」などがある。学術情報サービスという点からみると、独立行政法人科学技術振興機構が運営する「J-STAGE」、国立情報学研究所の「CiNii Articles」、Google 社の「Google Scholar」、Microsoft 社の「Microsoft Academic Search」、国立国会図書館の「NDL ONLINE」などが挙げられるし、個人が作成しているものも含めれば枚挙にいとまがない。

これらのデータベースは、当然すべて同じデータ項目を持っているわけではない。作成者の目的に応じて構築方法も違えば、検索方法や結果の表示方法、周辺情報へのアクセス等、各々違った特徴がある。利用者は、これらのデータベースを自身の目的に合わせて活用し、必要な情報を得るのだが、利用者の検索レベルは様々で、データベースの仕組みを熟知し、検索に長けているものばかりではない。中には適切な検索方法や検索ワードが分からず、求めるデータはあるのにそのデータを見つけ出せない場合もある。

データベースを構築する際に、全ての利用者の検索技術や使用目的を聞き取り調査して作りこむことは不可能である。だからといって、作ってから手直ししようとしても、一度稼働し始めたデータベースを停止させて作り直したり、大規模修正を行ったりということは難し

い。その結果、利用者にとって「使いにくい」と感じるデータベースとなってしまう。

しかし、Web ブラウザを使って利用できるデータベースの「使いやすさ」のうち、表示方法や操作方法については、ブラウザの拡張機能を使用することで、既存のデータベースを停止させることなく探ることが可能である。本研究では、2019年10月時点で64.92%のシェアを誇るブラウザ「Google Chrome」の拡張機能を使用して、既存のデータベースを用いながら人文科学系データベースの「使いやすさ」を探る方法を紹介すると同時に、どのような点に留意してデータベースの構築・運営を行うべきかを提示したい。

また、また、将来的にはスマートフォンやタブレット端末からの利用も想定すべきであるが、Chrome 拡張機能はモバイル版では使用できないため、本論では検証していない。

## 2. データベースの比較

本論では、国内では最大規模の蔵書データベースである「NDL ONLINE」、学術論文データベースとして規模の大きい「CiNii Articles」、人文系文献情報データベースとして規模の大きい「国文学論文目録データベース」、国文学論文目録データベースと同じ人間文化研究機構のデータベースである「日本語研究・日本語教育文献」データベースを取り上げる。

各データベースの概要と特徴は次の通りである。機能の比較については別表をご覧いただきたい。

### 2-1. NDL ONLINE

**概要:** 国立国会図書館検索・申込オンラインサービス。国立国会図書館(以下 NDL)の所蔵資料及び NDL で利用可能なデジタルコンテンツについて検索及び各種申し込みができる。NDL が収集した図書館資料及びデジタルコンテンツ、NDL が製作した学術文献録音図書主題情報(視聴覚資料)、主題情報等を検索の対象としている。

**特徴:** オンラインで資料の複写申請、閲覧申請を行うことができる。また、インターネットから利用者登録を行うことでマイリスト機能(選択した資料の書誌情報を保

存できる機能)を使うことができるようになる。



〈NDL ONLINE:キーワード『万葉集』での検索結果〉

## 2-2. CiNii Articles

**概要:**国立情報学研究所が管理・運営する、学術論文データベース。学協会刊行物・大学研究紀要・国立国会図書館の雑誌記事索引データベースなど、学術論文情報を検索の対象としている。

**特徴:**目録情報から本文にアクセスが可能。ただし、PDF化されており、かつ閲覧可能な状態でWeb上に存在しているものに限る。また、被引用文献の一覧を表示することができる。

連携しているデータベースから横断検索を行うことで膨大な検索結果を得ることが可能。ただし、全文検索のため、たとえば著者が同姓同名の別人であってもすべて結果として表示される。



〈CiNii Articles:キーワード『万葉集』での検索結果〉

## 2-3. 国文学論文目録データベース

**概要:**国文学研究資料館が管理・運営を行う、国文学関係の論文目録データベース。元々は『国文学年鑑』として出版されていた学術論文目録をWebでも検索が可能なデータベースとして作り替えたもの。現在では、国文学研究資料館で所蔵している、日本国内で発表された雑誌紀要単行本(論文集)等に収められ

た、日本文学・日本語学・日本語教育の研究論文に関する目録情報を検索の対象としている。

**特徴:**論文毎にキーワードが入力されており、タイトル、著者名だけでなく、キーワードが一致することで検索結果として表示される。ただし、キーワードは可視化されていない。

タイトルや発行年だけでなく、時代分類や分野を選択することでも検索が可能(図表)。

データが入力されていない部分(英文執筆等)は、空白表示になる。

国文学研究資料館 OPAC (Online Public Access Catalog) と連携しており、請求記号をクリックすることで同館の OPAC にアクセスが可能。



〈検索画面:時代分類/分野〉

## 2-4. 日本語研究・日本語教育文献データベース

**概要:**国立国語学研究所が管理・運営する論文目録データベース。『国語年鑑』『日本語教育年鑑』に掲載されている文献情報を元にしており、1950年から現在までの関係論文・図書を検索の対象としている。日本語学及び日本語教育に関する研究文献のデータベースとして運営されている。

**特徴:**文字コードの選択が可能(Shift-JISとUTF-8)。検索結果は表形式とリスト形式のいずれかで閲覧できる。表形式では登録されているキーワード、章タイトル、分類分野が表示されるほか、機関リポジトリなどがあるデータに関しては、「本文表示」のアイコンをクリックすることで本文へのアクセスが可能。

日本語研究・日本語教育文献データベースの表形式、リスト形式それぞれの検索結果画面に関しては、後ほど詳しく紹介する。

《データベース比較表》

データベース 名称	NDL ONLINE	CiNii Articles	国文学論文目録 データベース	日本語研究・ 日本語教育文献 データベース
表記揺れ(*1)	○	○	○	×
文字のハイライト	○	○	○(*2)	○
検索結果の ダウンロード	Tsv BibTeX MARC(*3)	RefWorks EndNote Mendeley (*4)	Excel, CSV TSN, JSON (*5)	Excel, CSV, BibTex (*6)
検索結果の 表示形式	リスト形式	リスト形式、 Refer/BibIX, RIS、 BibTex, TSV	リスト形式 (一覧表示) リスト形式(一覧表 示+ 詳細表示)	リスト形式 表形式
検索結果背景色	×	奇数行と偶数行で 色分け	奇数行と偶数行で 色分け	奇数行と偶数行で 色分け(*7)
並べ替え	タイトル 50 音順、 著者名 50 音順、 出版年、請求記号	出版年、引用件数 タイトル 50 音順、 刊行物名 50 音順、	ID、論文執筆者、 西暦年	×
検索結果表示件数 (単位:件)	20、50、100	20、50、100、200	50、100、200	20、50、100

\*1 「万葉集」と「萬葉集」/「芥川龍之介」と「芥川竜之介」等

\*2 検索結果に表示されないキーワードがあるため、検索結果としてはデータが表示されるが、ハイライト箇所がない場合もある

\*3 MARC 形式でダウンロードを行う場合は、NDL-Bib でのログインが必要

\*4 各文献管理サービスに直接書き出しが可能

\*5 検索結果が 1,500 件程度までのものに限る

\*6 表形式の場合に限る

\*7 表形式での表示に限る

### 3. Google Chrome とその拡張機能について

「Google Chrome(グーグル・クローム)」とは、Google 社が開発している Web ブラウザのことで、2008 年に公開された。Windows、MacOS、Linux、Android、iOS といった主要なオペレーティングシステムに対応しており、6 週間ごとに安定版のリリースが行われている。

手軽に「拡張機能」と呼ばれる機能を増やしたり強化したりするプログラム(スクリプトと呼ばれる)を開発・公開・導入することができるため、ブラウザそのものには備わっていない機能でも、必要に応じて容易に取得・使用することができる。

この「拡張機能」は、JavaScript や HTML、CSS とい

った Web の技術を使って開発するため、Java や C#、Swift といった一般的なプログラミング開発に比べると開発のための敷居が低い。このため、Chrome ウェブストアには多数の拡張機能が公開されている。

本研究では「表示された画面を操作する」ことに主眼を置くが、それだけでなく、右クリックで表示されるメニューを増やしたり、特定の Web サービスと連携して便利に使用できたりする拡張機能もあることを申し添えておきたい。

#### 4. データベースにおけるユーザビリティ

データベースにおける検索結果の表示方法は、大きく分けるとリスト形式と表形式の二つがある。

リスト形式の表示方法は「行」のみを持つ。場合によっては複数の項目を持つこともあるが、項目数と一覧

性は基本的に反比例の関係にある。よって、データ数のみを知りたい場合や、検索結果の量が少ない場合などは便利だが、「項目 A」に”o”と入っているデータを探したい、という使い方には不向きである。

表形式の表示方法は「行」に加えて複数の「項目」を持ち、項目数が増えても基本的に一覧性は下がらない。ただし、横スクロールが必要になるほど多くの項目を表示した場合は、一覧性は下がる。

データベースの機能として、表形式とリスト形式をもつ日本語研究・日本語教育文献データベースを例にとると、表形式での表示は、項目が多くても視認性が高くなり、多くの情報を一度に見ることが出来る。それに比べてリスト形式では視認性を高くしようとすると項目数をしぼる必要があり、逆に項目数を増やそうとすると行数が増えるため視認性が低くなる。



〈日本語研究・日本語教育文献データベース:リスト形式での表示〉



〈日本語研究・日本語教育文献データベース:表形式での表示〉

このように、表形式は一覧性が高いものの、モニターの画面サイズがあまり大きくない場合に、項目が折り返されてしまうため、かえって見にくくなることがあることが

分かる。では、リスト形式でも項目数と視認性を両立させる方法として、「タイトルなど主要な項目以外は折りたたみ、

必要な時だけ表示する」という方法はどうか。  
 Google Chromeの機能拡張を用いて、実際に検証してみた。

まず、〈図 1-1〉は、国文学論文目録データベースで「万葉集」を簡易検索した結果画面である。

リスト形式で表示されているため、タイトルだけをざっと見る、あるいはデータの総数やどのような傾向があるかを検索するような場合には便利だが、それがいつ(出版年月日)、どういう雑誌の(誌著名)、何号(巻号/通巻)に掲載されているものかを詳しく見ていくことは困難である。もちろん、詳細検索画面で〈出版年〉、〈掲載誌名〉、等で絞り込むことは可能だが、使い方は研究者によって異なるため、一つの方法が絶対正しいとは言えない。また、元々が空欄の項目(執筆者名(よみ)、英文執筆等)はカンマ区切りで空欄になっている

ため、見づらい部分もある。

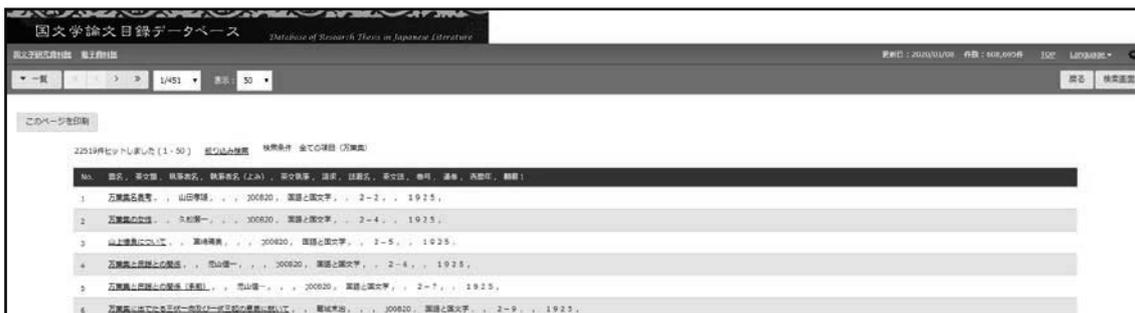
次に、〈図 1-2〉は、同じくキーワード「万葉集」で検索を行い、リスト形式+詳細画面で表示させたものである。

左側にリスト形式の画面、右側に詳細画面が表示されている。検索結果が少なければ、このような詳細画面を用いて検索することも難しくないが、膨大な量のデータがある場合、データひとつひとつを表示してチェックしていく作業は時間も手間もかかってしまう。

そこで、Google Chrome 拡張機能を用いてリスト形式で折りたためるようにした結果が、〈図 2-1〉及び〈図 2-2〉である。

最初は〈図 2-1〉のように表示されるため一覧性が高いが、「▼」をクリックすると、隠されていた項目が表示され、多くの項目を見ることもできる(図 2-2)。

また、表形式で表示させたものが〈図 3〉の通りとなる。



〈図 1-1〉: 検索結果一覧画面



〈図 1-1〉: 検索結果一覧+詳細画面



〔図 2-1〕



〔図 2-2〕



〔図 3〕

ここで注意しないといけないのは、リスト形式であっても表形式であっても、データベースとして表に出している項目しか表示させることができない点である。国文学論文目録データベースでは、可視化されていない情報として、論文毎に論文の中で取り上げられている作品名や作家名などが存在する。こういったデータは表には出していないため、拡張機能を使って見やすく表示しようとしても表示させることは出来ない。

データベースを作成する側は当然、「必要とされている情報だろう」と考える項目を表示しているのだが、どのような情報が必要でどのような情報が不要かは利用者により異なる。それならば、入力されている項目は「理由があって見せたくないもの」以外はすべて表示されるようにし、利用者が必要に応じて表示項目を選択できるようにすることで、より汎用性の高いデータベースとして利用してもらうことが出来るのではないだろうか。

ただし、表示されていなくともデータ通信量は増えるため、通信量に制限がある、あるいは通信回線が遅い利用者への配慮は必要である。

## 5. まとめ

ブラウザの拡張機能を使用することで、「データベースを利用者が使いやすい形に変えることができる」ということは、今まで述べてきたとおりである。しかし、使用するブラウザは利用者によって異なり、必ず Google Chrome を利用しているとは限らない。また、拡張機能の開発には技術と労力が必要になるため、研究者が手軽に開発することは現実的ではなく、システム会社等に外注する必要があるだろう。さらに、Chrome の拡張機能は現状スマートフォン版には対応していないため、増えつつあるスマートフォンユーザーへの対応は難しく、「データベースの利用者に拡張機能を使ってもらおう」という解決策は現実的ではない。

もし、拡張機能が必要なデータベースと、拡張機能が不要なデータベースがある場合、利用者が拡張機能の不要なデータベースを選ぶのは自然なことである。データベースは格納されている「情報」が命であることは言うまでもないが、どんなに良い情報を持ったデータベースでも、利用されなければ「存在価値がない」と評価されても仕方がない。

つまり、本論文で示した「使いやすさ」は、実際には各データベースがそれぞれに実践する必要があるといえる。大規模な構築や改修の計画がなくても、紹介した方法を実践して、日々「データベースのユーザビリティ」について検証を行うことが必要である。そうした検証の積み重ねは、新しいデータベースの構築や改修をより良くするために欠かせない。

また、それに加えて利用者の声に耳を傾けることも大切である。利用者のちょっとした違和感、こうだったらいいのにという思い、そういった声を得るための努力もまた、日々行うことが必要である。

## 6. 謝辞

本論文を作成するにあたり、国文学研究資料館論文データベース室のご協力を頂いた。ここに感謝の意

を記す。

## 7. 参考文献

- [1] 「人文系データベース」における相互運用性をめぐる諸問題」じんもんこん 2008 論文集
- [2] Web ブラウザシェアランキング TOP10(日本国内・世界)| ソフトウェアテスト・第三者検証ならウェブレッジ  
[https://webrage.jp/techblog/pc\\_browser\\_share/](https://webrage.jp/techblog/pc_browser_share/)

## 8. 附録

- ・ A New Version Of Google Chrome Now Due Every Six Weeks  
<https://techcrunch.com/2010/07/22/google-chrome-versions/>
- ・ ND L ONLINE (国立国会図書館検索・申込オンラインサービス)  
<https://ndlonline.ndl.go.jp/>
- ・ CiNii Articles  
<https://ci.nii.ac.jp/>
- ・ 国文学論文目録データベース  
<https://base1.nijl.ac.jp/~rombun/>
- ・ 日本語研究・日本語教育文献データベース  
<https://bibdb.ninjal.ac.jp/bunken/ja/>
- ・ アーカイブズ学文献データベース  
<http://base1.nijl.ac.jp/~archiadodb/>
- ・ Web データに基づく用例データベース  
<https://www2.ninjal.ac.jp/lrc/index.php?Web%A5%7%A1%BC%A5%BF%A4%CB%B4%F0%A4%C5%A4%AF%CD%D1%CE%E3%A5%C7%A1%BC%A5%BF%A5%D9%A1%BC%A5%B9>
- ・ J-STAGE  
<https://www.jstage.jst.go.jp/browse/-char/ja>
- ・ Google Scholar  
<https://scholar.google.co.jp/>
- ・ Microsoft Academic Search  
<https://academic.microsoft.com/home>
- ・ 国文学研究資料館 OPAC  
<https://opac.nijl.ac.jp/>

# 日本語学習ディクテーション・データベースの開発 Development of database of dictation practice for Japanese learners

山元 啓史  
Hilofumi Yamamoto

ホドシチェク ボル  
Hodošček Bor

東京工業大学 環境社会理工学院 東京都目黒区大岡山 2-12-1  
Tokyo Institute of Technology, 2-12-1, Ookayama Meguro Tokyo Japan 152-8850  
大阪大学大学院 言語文化研究科 大阪府豊中市待兼山町 1-8  
Osaka University, 1-8, Machikaneyamacho Toyonaka Osaka Japan 560-0043

**概要:** 東京工業大学で日本語を学ぶ留学生のためにディクテーション・データベースを開発している。このデータベースは、まったく日本語の前提知識がない学習者を対象に、日本語の発音・語順・活用・表現をディクテーションを通して、学ぶことを目的としている。短い文の聞き取りからはじめ、次第に長い文・複雑な文を学ぶ。学習者は文を聞いて（あるいは動画を見て）、セリフの一部の空欄部をキーボードからローマ字でタイプする。開発データは日常日本語口語表現 1500 文である。週 5 日 1 日 10 文 150 日（3 単位）分を開発した。すべての文には英語訳をつけ、文の意味はわかるようにした。文はできるだけ滑らかに話せる、自然な口語表現に限定した。また、日常生活に用いられる俗語・固有名詞等を意図的に多く取り入れた。開発したデータベースを利用して、オンライン・ディクテーション練習システムを作成し、日本語初級学習者に試用してもらった。本研究発表では、データベースの開発とディクテーション練習システムの試用アンケートから得られた結果について報告する。

**Abstract:** We have been developing a dictation database for international students learning Japanese at the Tokyo Institute of Technology. This database is for students who do not have any preexisting knowledge of the Japanese language. It provides learners basic knowledge through dictation practices, such as pronunciation, word order, verb conjugations, and daily expressions. Learners listen to relatively short sentences first, after which the length is gradually extended and complexity increased. On the practice site, learners fill in the blank part of the dialogue by typing on the keyboard while listening to the sentence or watching a video. We have developed 1500 Japanese colloquial sentences, which span 150 days (3 credits) of content corresponding to a 5 days a week, 10 sentences per day class. All the sentences are accompanied with English translations, so that learners can understand the meaning of each sentence and the situation. The sentences are limited to colloquial expressions which allow learners to speak smoothly and naturally, and the database deliberately includes slang words and proper names frequently used in daily life. In this study, we will address the development of the database, and will also mention the impressions of students who used the database through the online practice system.

**キーワード:** 日本語学習、会話、口頭表現、リスニング、俗語、語順

**Keywords:** Japanese language learning, colloquial expressions, listening, slang, word order

## 1 はじめに

本稿では初級学習者を対象に口頭表現 (Colloquial Japanese) の習得を目的としたディクテーション用データベースの開発について報告する。

「こういうことだったんですか」とは来日しばらくの後に筆者がクラスを担当した学生から発せられたことばであった。日本語の教科書通りに忠実に日本語を学んだが、来日して日本で触れた日本語は語順が違う、単語が違う、略語も俗語も多いことに気づいた。学生は「本当 (現実通りの日本語) を教えてもらっていたら、もっと早くうまくなって、ここまで苦勞することはなかった」という。この発言から、教科書や教室で教える内容は「事実即していない」ということがわかる。この問題解決の一助として、ディクテーションの訓練を取り上げる。

### 1.1 ディクテーションの訓練領域

ディクテーションは聴解と受け取られるかも知れないが、聴解能力というよりも、同じ文を何回も聞いて何であるかを感じ取る能力である。日常の聴解行動は一回限りであるため、ディクテーションのように何回も聞き取る行動とは別ものである。ディクテーションは手続き的知識、処理の自動化 (杉浦他 2002: 106) に貢献するものと考えられている<sup>\*1</sup>。ただし、教育の現場では知識の手続化に必要な練習が絶対的に少ないことが指摘されている (杉浦他 2002: 106)<sup>\*2</sup>。

ディクテーションで使われる文は、自然なものが望ましい。自然な言語インプットから言語要素の学習理論があり (Krashen and Terrell 1983)、自然な言語インプットに習得効果があることがわかっている (Blau 1990: 746)。すなわち、一息で言えない文や書きことば、スピーチの台本、ニュース番組・ナ

レーションなどの話の専門家が読み上げるもの、教科書に記載された材料などもふさわしくない。残念ながら、中級向け、上級向けの英語ディクテーションサイトは多いが、初級向けのは多くない。現実の言語音声を書き取るのは初級の学習者には難しいと考えられているようで、初級教科書の例文や CD には語彙の調整・加工などが見られる。初級は基本的な語彙、文型・文法を学び、中級以上から実際の材料を使った教育を行い、段階的に指導する意図であろうが、事実即した自然な材料が提供されているかどうかはわからない。

### 1.2 ディクテーションの利点と要件

ディクテーションは、機能的な語学力の比較的シンプルで信頼性が高いテストとして注目されている (Savignon 1982)。単なる綴り字テストとは異なり、ノートテイキングのように学習者が日常生活で行うタスクと似たものと考えられている (Cohen 1980)。また、他の語学力テストとの間に 0.96 で相関があることがわかっており (Oller and Streiff 1975)、言語能力の正確かつ効率的な尺度であることが示されている (Cai 2013, Oller and Streiff 1975)。一方で、文が易しい場合は全文を聞かずとも答えられるため単語認知の能力しか測れない (Buck 2001) といった負の評価もあり、初級学習者を意図して、文をやさしく改変する必要もないことが想像できる。ディクテーションの利点は多肢選択式の内容理解問題などに比べて作成が容易である。文構造についても過剰に考慮する必要もないと言われる (Blau 1990: 752)。

ディクテーションは認知処理の統合であり (Oller 1971)、単語の認知でさえも文脈から切り離して処理することができないという (Frauenfelder and Tyler 1987)。話しことばは短い間に多くの異なる音がすばやく連続的に発せられるもので、即時的で効率的で柔軟な談話 (Nishimura 2010: 72) である。それゆえ話しことばを模したのではなく、自然そのものの特性が不可欠であることが予想される。その自然さがあれば、文脈中から文法を潜在的に学ぶ効果も示唆されている (Oller and Streiff 1975)。

\*1 リスニングにおける問題は、音韻知識が不足していること、音韻知識と語彙・表現などの言語情報との結び付きを自動的に処理できるか、聞いて即理解する能力が運用できるかどうかであり、これを可能にする処理を養成する必要があると言われている (杉浦他 2002: 106)。

\*2 手続化: 知識を行為に変換 (手続化) し、複数の手続を組み合わせて一連の行為連鎖を自動的に行えるように何度も練習する必要があるといわれている (Gagné et al. 1993)。

## 2 方法

### 2.1 センテンスの特徴と作り方

データベースに収録したセンテンスの特徴はすべて口語文であることである。語順は、因習的に考えられている主語・目的語・述語のような紋切り型の語順に因われず、むしろ倒置や頭に思いついた順の語順を積極的に取り上げる。話者間で互い既知の情報は発話されず、語の省略が行われるが、この特徴もデータベースに収録し、省略に相当する意味は英訳あるいは映像で補完する。言いよどみ、言い直し、フラグメント（途中で切れの文や句）なども収録する。単語は、固有名詞を積極的に使用する。特に地名、個人名、商品名などを取り入れる。長い文を避け、3文節ないし5文節で表すものに限定するか、対話の形式にして、できるだけ短い文で収録する。毎日続けることを主眼とし、同じことを同じように繰り返すより、似てはいるが日々内容的に違うことが継続できるよう配慮する。

### 2.2 データの形式・内容構成・例

回答項目1つを1レコードとする。1レコードは、1)レコードID、2)発話文、3)回答箇所を空欄としたカッコ付の発話文、4)正答文字列、5)対訳英文文字列、の5つのフィールドからなる。発話文は自然な口語である。単純化した発話を提示する(Leow 2003)。短い発話で文だけでは理解しにくいいため、英語訳がついている。各レコードには発話文に相当する状況を示す映像をつける予定である。俗語(マジっすか)、固有名詞(スタバ、ファミマ)、縮約形(てしまう→ちゃう)、言い換えのための接続詞(つまり、つーか、で…)、言いよどみ(ああ、まあ、そのう)等はできるだけ書きことばで一般的に記される形を使用する。データベースで扱われる内容は表1に示されるトピックである。ディクテーション練習の例を表2に示す。

## 3 試用結果

ディクテーションデータベースの試用は、東京工業大学日本語・日本文化科目「Strategic Japanese

(以下、TOJ51: Topic Oriented Japanese 51)」の宿題として、7週実施した。TOJ51は12名(一時的に15名)受講者で、2019年4月9日から6月6日の7週間行われた。受講者は全員そのクォータ前のクォータに来日した日本語学習経験がほとんどないか、初級前半のレベルで、学習者によっては日本語の音や文字も知らない。クラスの目的は、日本語学習ではなく、日本文化学習の一環として、日本語の学び方について互いにディスカッションを行いながら、知識を深める文化学習クラスである。授業は、すべて英語で行われ、一部、日本語を使ったペア練習が盛り込まれてはいるが、それは主目的ではない。授業スケジュールと授業概要は表3のとおりである。

### 3.1 実施手順

毎回の実施は、ログインからはじまり、ディクテーション文を聞く、聞こえた文をタイプする、答え合わせ(自動)を単純に繰り返す。1週間に70文、8週間計500文のディクテーションを行い、1500文のうちの1単位分のみを実施した。練習後の質問として、「むずかしかったディクテーションはどれだったか」、「うまくできたディクテーションはどれだったか」を尋ねたが、これは必須とはしなかった。

### 3.2 アンケート結果

各練習実施後6回にわたって、ディクテーションについてのアンケートを行った。アンケートは1)「継続しているかどうか」、2)「難しいかどうか」を毎回聞いた。それ以外に、コース前半時期には3)「ディクテーションが好きか」、ディクテーション実施後、5)「ディクテーションについてどう思っているか」について聞いた。コース中盤から後半時期には4)「ディクテーションによって日本語に関する知識は増えたか」、6)「その他コメント」を尋ねた。1-4)の項目については表4に示す。

「毎日続けたか」という質問には、ごく一部のやり忘れの学生を除けば、ほぼ全員が毎日実施している(実施ロードの問題)。「難しいか」という質問には、多くが難しいと答えている。「好きか」という質問には、おおむね「好き」である。「知識は増え

表 1: 日本語ディクテーションデータベース収録内容

Volume 1	Volume 2	Volume 3
<b>Week 1</b> あいさつことば 1) あいさつ 2) 招待する・される 3) 食べ物 4) 健康と病気 5) ほめる 6) イ形容詞 7) ナ形容詞	<b>Week 1</b> コミュニケーションのための文 1 1) 会話で役に立つ表現 2) 曖昧な表現だが、必要な表現 3) 好みと味 4) 勝ち負け 5) やりすぎ/軽く見る/過小評価する 6) 敬語と感謝 7) 健康と病気	<b>Week 1</b> さあ、はじめましょう 1) 報告書 2) 仕事を終えて 3) 散歩に出かけて 4) 旅行 5) 文化の日 6) わかる・わからない 7) 心機一転
<b>Week 2</b> 好み 8) 好きと嫌い 9) よいとわるい 10) 好みを尋ねる 11) 形容詞 12) 願望 13) ほめる 14) 何かする	<b>Week 2</b> ストラテジックな文 1 8) 手伝う・役に立つ 9) 何かに気づく 10) 易しさと難しさ 11) ちょっと前/たった今 12) 驚きと発見 13) いつでも言える 14) 一応・とりあえず	<b>Week 2</b> 日常のできごと 8) 味と風味 9) お菓子 10) 仕事を終わらせる 11) 困難なこと 12) 今週の予定 13) 事件 14) 駅のアナウンス
<b>Week 3</b> 食べ物 15) 成分 16) 味が良い・悪い 17) 調理法 1 18) 調理法 2 19) 食べ物とレストラン 1 20) 食べ物とレストラン 2 21) 食べ物とレストラン 3	<b>Week 3</b> コミュニケーションのための文 2 15) アドバイス・進言・お勧め 16) 味覚と複文 17) やり直す・作り直す・訂正する 18) 材料 19) 特別に/格別に 20) 経験 21) 希望/期待	<b>Week 3</b> 健康になろう 15) 病気と保険 16) 筋肉痛と肩こり 17) 病気の症状 18) 食事と健康 19) 歯医者 20) 病気と病院 21) カロリーとダイエット
<b>Week 4</b> 遊びの文化 22) じゃんけん 23) ゲーム・スポーツ 24) 花見 25) ポップカルチャー 26) 福笑い 1 27) 福笑い 2	<b>Week 4</b> ストラテジックな文 2 22) できるだけ、好きだけ 23) 比較する 24) 季節 25) 人気がある 26) 技能がある/上手にできる 27) ようだ/そうだ 28) 間違える	<b>Week 4</b> 半分終わりました 22) 天気のいい日に 23) 池のほとりを散歩 21) 訪問 25) レストランの予約 26) ヘアカット 27) 外出 28) エレベーターとエスカレーター
<b>Week 5</b> インタラクションとストラテジ 28) 何かに挑戦する・しない 29) 後悔する 30) 鑑賞する 31) 示唆する 32) 経験する 33) 確認 34) ストラテジーの表現 1 35) ストラテジーの表現 2	<b>Week 5</b> 日本語らしい日本語 29) 体に関する慣用句 30) 通用する/実際に使える 31) 親族呼称 32) 同じ/似ている 33) 敬語表現 34) 動物を使った慣用句 35) 四字熟語	<b>Week 5</b> 毎日の過ごし方 29) 授業登録 30) 歯ブラシ・歯磨き 31) 締め切りが迫っている時 32) 気分が優れない時 33) 微妙なやり取り 34) パスポートの申請 35) 免許証の更新
<b>Week 6</b> ストラテジーの表現 36) ストラテジーの表現 3 37) ストラテジーの表現 4 38) ストラテジーの表現 5 39) ストラテジーの表現 6 40) ストラテジーの表現 7 41) 補語 1 42) 補語 2	<b>Week 6</b> 動詞関連 36) 使役 37) 副詞と動詞 38) 複合動詞 39) 自動詞と他動詞 40) 受身形 41) 命令形 42) 意向形	<b>Week 6</b> 未来の出来事 36) 見たことない 37) 移動 38) イントクイズ 39) 買い物 40) 明日の予定 41) 誕生日 42) 小学校訪問
<b>Week 7</b> 自然な話し方 43) 対人関係表現 1 44) 対人関係表現 2 45) 対人関係表現 3 46) 対人関係表現 4 47) 対人関係表現 5 48) 対人関係表現 6 49) 対人関係表現 7	<b>Week 7</b> 複文 43) 同時に/てから 44) やりもらい 45) 条件節 46) 完了表現 47) 名詞修飾 48) 擬音語・擬態語 49) 可能形	<b>Week 7</b> いよいよ最終週 43) 研究倫理 44) 人生に必要なこと 45) タワーゲーム 46) 洗濯物 47) 居酒屋 48) さあやるぞ 49) 学園祭
<b>Week 8</b> まとめ 50) Summary	<b>Week 8</b> まとめ 50) 挿入句	<b>Week 8</b> まとめ 50) 最終試験

たか」という質問には、ほとんどが増えたと答えており、おおむね良好な結果が得られている。

#### 4 考察

試行で得られたコメントを元に、目的のデータベースに近づいているかどうかを検討する。

まず「ディクテーションは難しいか(表5)」からは、多くが、むずかしいことは認めていることがわかる。初級学習者にとって自然な発話は速く、その速さに合わせて、正確にタイプして綴る、長音や濁音の音の認識、文の区切りの認識、新単語の認識な

どで苦勞しており、文の全体像の把握、複雑さの処理を問題として認識していることがわかる。また、過剰にできないときにはいやになることも事実だろう。一方で、徐々に簡単になってきたと感じる、自分で調べて解決する、上手になってきておもしろさを感じる、など難しいながらも克服していることもわかる。

つぎに「ディクテーションは好きか(表6)」からは、短い時間に集中的に練習できる、毎日の練習が習慣化している、毎日日本語に触れられる点をあげていることがわかる。一方で、なかなか正解にならないから好きになれない、という発言から、フィー

表 2: ディクテーション「何かやるか、やらないか」の例; 部分聞き取りの場合にはアンダーライン部のみを書き取る。

1. A: <u>トランプ</u> 、しよ?	Let's play cards!
B: <u>いやだ!</u>	I don't wanna play it.
2. A: どうして?	Why?
B: まけるの、 <u>きらい</u> 。	I hate losing.
3. A: <u>とにかく</u> やってみよ。	Let's do it anyway.
B: うーん。	Well, no...
4. A: <u>やってみたら</u> どう?	Why don't you try it?
B: そう?	You think so?
5. A: パチンコ、 <u>よく</u> しますね。	I often play pachinko.
B: <u>そう</u> ですね。毎日ですね。	Yes, you play it everyday.
6. A: マージャンもときどきします。	I also do mah-jong from time to time.
B: <u>本当</u> は毎日ですね。	In fact, you play it everyday.
7. A: <u>でも</u> 、 <u>付き合</u> い程度ですね。	But, I am a kind of social-player.
B: <u>でも</u> 、毎日ですね。	But, you play it everyday.
8. A: <u>運動</u> もしていますか。	Are you doing exercises?
B: <u>ラジオ</u> 体操ぐらいですね。	Only radio exercises.
9. A: <u>あまり</u> 運動、 <u>しな</u> いんですか。	Not much exercises?
B: <u>運動</u> らしい運動はしませんね、 <u>ほとんど</u> 。	I seldomly do exercise kind a thing.
10. A: <u>マラソン</u> 、 <u>はじ</u> めませんか。	How about starting a marathon?!
B: <u>死</u> んでしまいますよ。	I would die.

表 3: 同時並行に行われた授業の内容; 2019 年度東京工業大学第 1 クォータ

授業回	日付	テーマ
1	4/09	学習ストラテジー: 音の捉え方
2	4/16	リスニング 1: 日本の歌
3	4/18	リスニング 2: 日本の歌
4	5/07	ビデオを見る 1: コメディ
5	5/16	ビデオを見る 2: ドラマ
6	5/21	文化とは 1: 日本人の行動
7	5/23	文化とは 2: 食べ物

ドバック（できていない点の明示的な指摘）を行う必要もあるだろう。

ディクテーション活動自体の認識「ディクテーションをどう考えているか（表 7）」については、発音に慣れるのに役立つ、多くの文は短いことが体験的に理解できている、濁音・清音の区別、促音・長音の認識が難しいことについて気づいていることがわかる。これらは、教師が大切だと言っても意味が

ほとんどない点であり、自分で気づいていることは良いことである。聞き取りよりもむしろ文の集中すべき練習であることに気づいている点も体験しなければ身につかない要素であろう。ただし、ローマ字で入力し、ローマ字文字列の一致によって正誤判定を行っていることが正解に至らなく、イライラ感が募ったことが想像できる。確かにタイプをするのは面倒であるが、自分の誤りを目視で確認できるこ

表 4: ディクテーション練習実施後の学生の意見 (数字は日付と人数)

授業実施回	2		3		4		5		6		7	
	4/16		4/18		5/7		5/16		5/21		5/23	
	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
1. 毎日続けているか	12	0	12	3	12	0	9	3	12	0	11	1
2. むずかしいか	8	4	6	8	11	1	9	3	9	3	7	5
3. 好きか	11	1	8	6	-	-	-	-	-	-	-	-
4. 知識は増えたか	-	-	-	-	12	0	11	1	12	0	12	0

とは利点である (飯野 2013: 79)。今後は、ローマ字入力で内部でカナ変換するか、カナ入力にするかして、内部のカナ文字列と照合し、正誤判定を行うようにするなどの改善が必要になる。

「ディクテーションによって知識は増えたか (表 8)」については、単語力、文の構造など言語要素に対する知識や認識だけでなく、徐々にわかってきた、日本語の知識が広がった、簡単な単語が違った音だったことに気がついた、長い時間日本語に触れることができた、単語と発音になれてきた、などの手続き的 (連続的・総合的) な知識の習得がおこっていることがわかった。

「その他コメント (表 9)」にはその他雑多な意見が見られる。注目すべきは、ディクテーションは自習でやるべきことで先生から学ぶものではないもの、習慣づけができた、と学び方 (ストラテジ) に関する点である。また、楽しい、おもしろいといったように楽しみながらできる内容であったことがわかる。

本研究において開発したディクテーションによる活動においても「ディクテーションをしなさい」以外に特別なインストラクションを与えていないことから、聞こえた音を書き取るだけのシンプルな練習が行われ、毎日の習慣となっていたことがわかる。また、自然な話しことばで、聞いて書き取れるものであれば、多少難しいレベルであってもよいこともわかった。何度も聞いているうちに、自分ができない理由を考えて工夫をするようになる、この工夫をストラテジというが、繰り返し自分で練習する、何

度も聞き直していることもわかった\*<sup>3</sup>。今後は、はじめは本来のスピードで聞くが、その後は、自分でスピードを一文毎に変えて、何度も繰り返して聞くことができる (Zhao 1997) 機能も追加すると良いだろう。また、自分の書いたところと正解とを見直して、自分のできないところを認識できるよう、フィードバック情報を提示したり、練習後、グループでディクテーションのコツを話し合う活動も計画できればよいと考える。

## 5 おわりに

本稿では、日本語学習ディクテーション・データベースの開発とそれを利用した学習システムを紹介し、実施アンケートの結果を報告した。その結果、タイプする単純な回答行動ではあるが、手続き的な知識の習得も行われていること、ローマ字入力に対応する点についても改善が求められることがわかった。

## 引用文献

- Blau, Eileen K. (1990) "The Effect of Syntax, Speed, and Pauses on Listening Comprehension", *TESOL Quarterly*, Vol. 24, pp. 746-753, 12.
- Buck, Gary (2001) *Assessing Listening*, Cambridge Language Assessment: Cambridge University Press.
- Cai, Hongwen (2013) "Partial dictation as a measure of EFL listening proficiency: Evidence from confirmatory factor analysis", *Language Testing*, Vol. 30, pp. 177-199, 04.

\*<sup>3</sup> ストラテジは個人で異なるので、教師が概念的に説明したり、行動で明示的に教えることはできない。

表 5: ディクテーション練習実施後の学生の意見「ディクテーションは難しいか」

1. 中国人だからなのかよくわからないけれども、むずかしい。	むのは難しいだろう。	11. 難しくないが、ローマ字綴りで困る。
2. 映画を見たり、歌を聞いたりして「か」と「が」、「た」と「だ」が区別できず聞き取れない。	6. 速く話すと難しいが、以前よりも簡単になってきた。	12. 文が長いときは難しいと思う。
3. 聞くことはあまりむずかしくないが、正確にタイプするのはむずかしい。	7. 上手になってきた。ある日の聞き取りはとてもおもしろかった。	13. 答えを書くのに文を区切りながら、何回も聞き直すことがある。
4. 長い文の場合、ちょっと難しい。	8. 難しくないが、注意深く聞いて正しく綴ることが求められる。	14. 長い文、複雑な文は難しいと思うが、本で新しい単語を調べたら、できた。
5. 語彙知識があまりないと全体をつか	9. 難しくない。自分としてはふつう。	15. スピードが速いと複雑に感じる。
	10. 最近、一般的に簡単になってきた。	16. し/ち、す/つ、とう/と、が区別困難。
		17. 正しく綴れないときにはいやになる。

表 6: ディクテーション練習実施後の学生の意見「ディクテーションは好きか」

1. 促音を聞き取れず正しく書けず難しいが、たくさんのが学べた。	るが良い練習だ。	間がかからないところが良い。
2. 日本語上達のチャンスが得られたので、この練習は好きだ。	5. よいと思う。しかし、答えを見るまで正解に至らないと少しやる気が失せるところがある。たぶん選択肢になると少し楽しくなるかも。	7. 簡単に効率良い毎日の練習だ。
3. もちろん、集中的に練習でき、役立つ。		8. 毎日日本語に触られるので助かる。
4. 意味や構造がわかるまで時間は掛か	6. 簡単に聞き取り練習ができるし、時	9. どうして正解にならないのかわからないから、あまり好きではない。

表 7: ディクテーション練習実施後の学生の意見「ディクテーションをどう考えているか」

1. 日本語の発音に慣れるのに役立つ。	そして、自信を持って教室外でことばを使うのに良い方法である。	たり“tu” だったり、混乱する。カナで書くなら間違えないのに。
2. 答えをチェックするのに時間が掛かる。正しい答えを入れたのに、まだ間違っているの、どこが違うのか指摘してくれたら、助かる。	6. たとえば、「し」と「つ」のように発音を直すのは難しいことがわかった。	13. 文構造が理解でき、新語も学べた。
3. 日本語聴解技能が向上した。	7. 本当の日本語の文が出てくると良い。	14. 語相互の結びつきが聞き取りにくく「た/だ」「か/が」の区別が難しい。
4. 教科書で習ったよりも、多くの文は短いことがわかった。また、研究室の同僚が話すときに、以前より話ことばの日本語がわかるようになった。	8. 聞き取り練習は大変だがおもしろい。	15. 1日5文にしたらよいだろう。
5. 非常に多くの単語を学ぶ良い方法だ。	9. “ou” と “oo” の区別は難しい。	16. 実際の日本語の文が学べる。
	10. 日常使われる文がたくさん学べる。	17. これは聞き取りよりもむしろ文の意味に集中すべき練習だ。
	11. 「た」と「だ」の区別が難しいが、難しさはちょうどいい。	
	12. 簡単だが、「つ」の正解は“tsu” だ	

Cohen, Andrew D (1980) *Testing Language Ability in the Classroom*: ERIC.

Frauenfelder, Uli H. and Lorraine Komisarjevsky Tyler (1987) “The process of spoken word recognition: An introduction”, *Cognition*, Vol. 25, No. 1-2, pp. 1-20.

Gagné, E.D., C.W. Yekovich, and F.R. Yekovich (1993) *The Cognitive Psychology of School Learning*: Harper Collins College Publishers.

飯野厚 (2013) 「英語教育におけるリスニング、シャドーイング、ディクテーションの関係」, 『法政大学多摩論集』, 第 29 巻, 67-81, mar.

Krashen, Stephen D. and Tracy D. Terrell (1983) *The Natural Approach: Language Acquisition in the Classroom*, Hayward, California: Alemany Press.

Leow, Ronald (2003) “Simplification and Second Lan-

guage Acquisition”, *World Englishes*, Vol. 16, pp. 291-296, 01.

Nishimura, Takeshi (2010) “Hanashi kotoba no seibutugakuteki kiban (The biological basis of spoken language)”, in Hisakazu Hasegawa ed. *Gengo to seibutugaku (Language and biology)*, Tokyo, Japan: Asakura shoten, Chap. 4, pp. 70-96.

Oller, John and Virginia Streiff (1975) “Dictation: A Test of Grammar-Based Expectancies”, *English Language Teaching*, Vol. 30, pp. 25-35, 10.

Oller, John (1971) “Dictation as a device for testing foreign language proficiency”, *ELT Journal*, Vol. 25, pp. 254-259, June.

Savignon, Sandra J. (1982) “Dictation as a measure of communicative competence in French as a

表 8: ディクテーション練習実施後の学生の意見「ディクテーションによって知識は増えたか」

1. 後で単語を調べて単語力がついた。	とができる。	11. 語彙もよく使う表現も増えた。
2. どのくらい覚えているかコース終了後自分でもう一度テストする。	8. 聞き取りはすこし上達した。混乱していた音がわかった。カジュアルな言い方がわかった。もちろん、新しい単語も覚えた。よい聞き取りの学び方。単語量は増えた。	12. 基礎的な単語や文構造を知ってしまったら、この練習は役立つ。
3. 日本語の単語を広げることができたし、日常の会話練習のおかげで実際に日常で使えるようになった。	9. 聞き取り練習は新しい単語を学ぶことで日本語の知識が広がった。モデル文が日常会話に使えた。	13. 新しい語を学んだ。
4. 練習しているうちに知識は増えた。	10. 日本語の単語と発音に慣れてきた。	14. 教室で習った「私」「いくら」のような簡単な単語が違った音だったので、さらに理解するのに難しかった。
5. 徐々に単語や文がわかってきた。		15. 知らない単語がまだ含まれていたので、聴解力向上に役立つ。
6. 練習した文に慣れ、学んだ。		
7. できるだけ長い時間言語に触れるこ		

表 9: ディクテーション練習実施後の学生の意見「その他コメント」

1. ディクテーションは自習でやるべきで、先生から学ぶものではない。	5. 楽しい。ローマ字で書くのでどこでもできるし、ラップトップでも。	役立つと思う。
2. これで自信をよりつけた。	6. おもしろいディクテーションで今までこんな経験したことがない。	8. 文法説明も付いているとよりよい。
3. 新しいことばを学ぶのに最適。	7. 単語間のポーズがあれば単語の意味を理解し、新しい語彙を覚えるのに	9. どの単語も正しく綴るのは難しい。
4. 毎日の聞き取りで、新しい語を学ぶ習慣をつけることができた。		10. 文が速く、音節を認識するには繰り返し聞く必要あり。

second language<sup>1</sup>”, *Language Learning*, Vol. 32, No. 1, pp. 33–47.

杉浦正利・竹内彰子・馬場今日子 (2002) 「リスニング能力養成のための自律学習：ディクテーションの効果」, 『言語文化論集』, 第 23 卷, 第 2 号, 105–121, mar.

Zhao, Yong (1997) “The effects of listeners’ control of speech rate on second language comprehension”, *Applied Linguistics*, Vol. 18, pp. 49–68.

## **基調講演**

**司会:石川徹也(TRC-ADEAC(株))**



## 歴史知識における「時」の表現と構造

赤石 美奈

法政大学・情報科学部

本研究では、膨大な史料中に存在する様々な時間を表す表現の抽出と、時間の構造の視覚化により、偶発的、あるいは必然的に生じた様々な事象の連なりにより表現される文脈を浮き彫りにし、潜在文脈を顕在化する手法について考える。

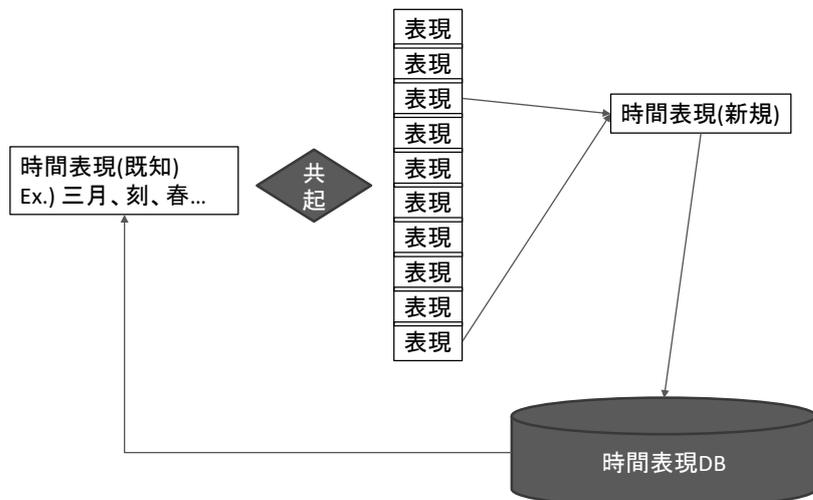
まず、「時間」概念の研究のためには、「時」を表す表現の抽出が必要である。様々な分野において、「時」を表す表現は、単なる暦や時間の表現のみではなく、自然現象の移り変わりや人間の心の変化などを表す表現なども含まれる。そこで、既知の時間表現を基にして関連語を抽出し、語の共起に基づく関係を視覚化するツールを利用し、新たな時間表現を見出す方法について述べる。

さらに、理論的に定義している時間構造を、実際のデータから見出すための手法について述べる。本研究においては、事象の変化の構造として、短期的な繰り返しの変化が循環し、再び以前と同じような状態に落ち着く変化構造や、長期的な視点で一定方向に変化し、その変化が常態となる変化構造に着目し、それらの構造(規則性)を抽出する手法について検討する。

# 歴史知識における 「時」の表現と構造

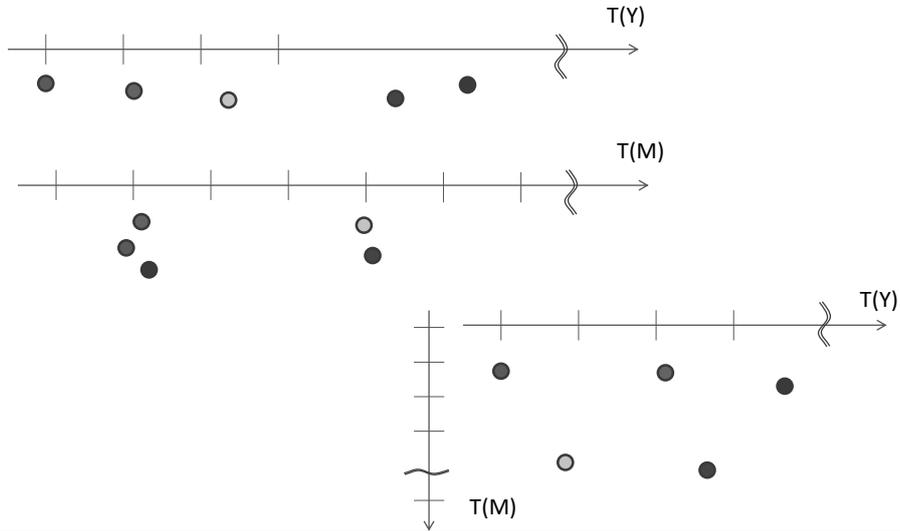
赤石 美奈 (法政大学・情報科学部)

## 時間表現抽出





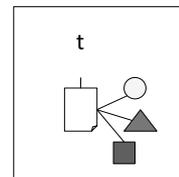
# 時間構造抽出



# 編年データ構造

- 編年データ構造

- テキスト
- タイムスタンプ(日付)
- メタデータ  
ex.) 人名, 場所, 事項 etc.



編年史料データベース D6選択 項目検索 結果

---

1/4190件

【編/冊】 大日本史料 1編2冊609頁

【和暦年月日】 昌泰1年10月20日(08960100200) 1条

【本文内容】 宇多上皇、近郊に御遊猟あらせられ、尋で、大和、河内、摂津等に御幸あらせらる。

【人名】 宇多、宇多上皇、宇多院

【官職】 上皇

【地名】 大和、河内、摂津

【事項】 御遊猟、遊猟、御幸、行幸、大和御幸、河内御幸、摂津御幸

【備考1】 01

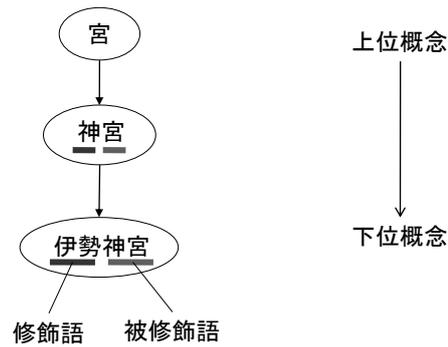
昌泰1年10月20日

宇多上皇、近郊に御遊猟あらせられ、.....

- 人名: 宇多上皇
- ▲ 地名: 大和、河内
- 事項: 御遊猟、御幸

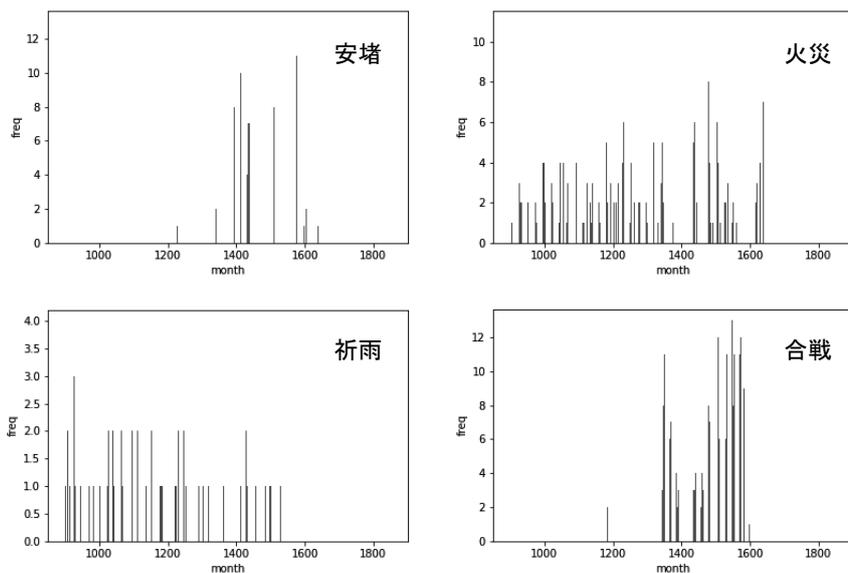
# 語彙オントロジー構築

- 概念階層の自動生成

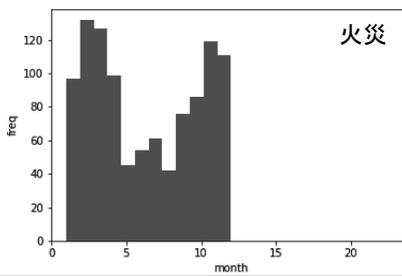
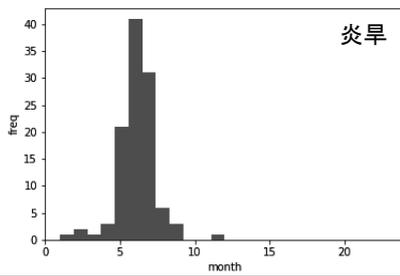
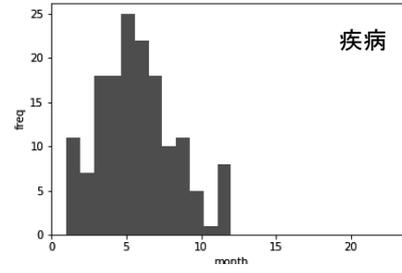
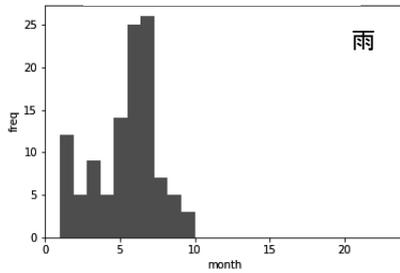


7

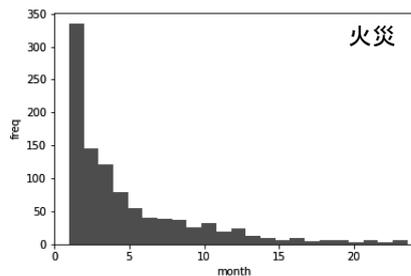
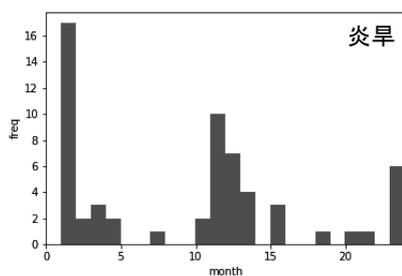
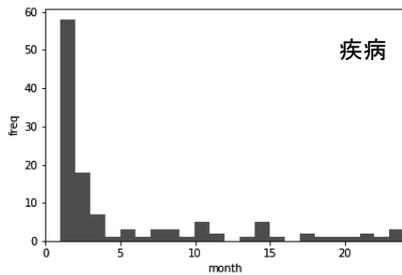
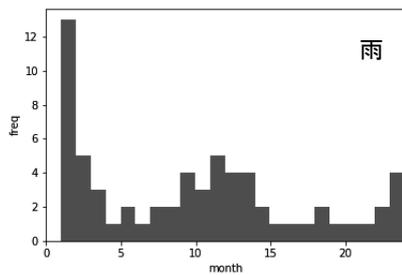
## 時代性: 事象出現頻度の年変化



## 季節性: 事象出現頻度の月変化



## 周期性: 事象出現間隔の月変化



# 概念?関連語(共起語)?の変化

百姓

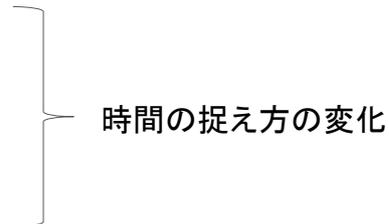
887-987	937-1037	987-1087	1037-1137	1087-1187	1137-1237	1187-1287	1237-1337	1287-1387	1337-1437	1387-1487	1437-1537	1487-1587	1537-1637	1587-1687
百姓	百姓	百姓	百姓	百姓	百姓	百姓	百姓	不輸	百姓	百姓	百姓	百姓	百姓	百姓
王臣家	陽明門	陽明門	改築	応徳	充行	住僧	留守所	百姓	荏胡麻	荏胡麻	名主	年貢	宛行	御法会
院宮	状	状	宸記	連歌	往	庁宣	庁宣	道場	逃散	逃散	年貢	名主	年貢	聖忌
諸家	愁訴	愁訴	故	連句	百首和歌	行幸御祈	断	邑	散在	散在	公事	住持	御祝	般舟三昧院
堰	水	召問	連句	三合	鷺	曲	東大寺領	制札	左京職	左京職	棄破	寺	役	見
失火	尾張国	尾張国	香椎	香椎	降雪	粉河寺	地頭	賊	不入	延暦寺	家	役	約	代官
禁	召問	勅旨駒牽	弟	三合厄	運上	地頭	住僧	沽却	延暦寺	地頭	住持	安堵	神主	総持寺
禁断	明	明	連歌	盆	屢	留守所	拜殿	人夫	守護使不入	守護使不入	遠乱	職	浅間社	鮭
田	乞巧奠	信濃勅旨駒牽	初雪	理	価	水田	録	落書	巷所	実相院	薙髮	宛行	社	興行

## まとめ

- 「時」の表現抽出



- 「時」の構造抽出





# 企業システム紹介



## デジタルアーカイブシステム ADEAC の紹介

田山健二

TRC-ADEAC 株式会社

デジタルアーカイブシステム ADEAC(A System of Digitalization and Exhibition for Archive Collection)は、多種多様な史資料等を高精細画像データとして、また必要に応じフルテキスト化し、メタデータと共に公開するシステムです。

ADEAC は、現在ジャパンサーチ(試験版)および EBSCO Discovery Service との連携システムとしても利用されています。

2020年2月1日現在の実績は以下のとおりです。

搭載機関: 101 機関

メタデータ(目録): 78,208 件

画像データ: 35,442 件

本文テキストデータ: 85,303 件

### 【公開事例】

- ・ 札幌市中央図書館／新札幌市史デジタルアーカイブ
- ・ 浜松市立中央図書館／浜松市文化遺産デジタルアーカイブ
- ・ 同志社大学・京都府立図書館共同事業／京都府立図書館貴重書コレクション
- ・ 日本ラグビーフットボール協会／日本ラグビー デジタルミュージアム

### 「WEB 版 デジタル伊能図」(パンフレットから抜粋)

「WEB 版デジタル伊能図」は、株式会社河出書房新社と東京カートグラフィック株式会社が共同制作した「デジタル伊能図」(DVD 版)を元に、伊能忠敬没後 200 年を記念してインターネットを介し利用できるようにしたものです。

伊能忠敬が全国を測量して作成した「伊能図」を元に測線、地名、測量日記、そして地図画像そのものを GIS で利用できるようにデータ化。「伊能図」と国土地理院の標準地図や空撮写真などとの重ね合わせを実現させました。パソコンやタブレットなどで、200 年前の地図とともにわが町の歴史散歩が可能となります。また「伊能図」と「測量日記」を連動させ、検索結果から日記と地図を合わせ見ることが可能です。

公共図書館・大学・研究機関 1 ライセンス ¥60,000 (年額・税別)

個人・研究会 1 ライセンス ¥12,000 (年額・税別)

# 「伊能図」の測量技術

太田武和

測量士補、元伊能忠敬記念館 非常勤説明員

## 1. 地図とは

地図とは、地表の諸物体・現象を一定の約束に従って縮尺し、記号・文字を用いて平面上に表現した図であり地形図・天気図などがあります。

地表の諸物体とは、私達が日常見ている土地の起伏、形態、水系、地表などのことを言い、地形図は、これらの諸物体（または地物）を、各地域の土地に分布する配置などを描いた図であり、天気図は、各地点での気圧、気温、雲の量、風向、風速など空の状態を表した図であります。

## 2. 伊能忠敬が作成した地図

伊能忠敬が作成した地図は地形図で日本列島の輪郭を表しています。縮尺違いの3つの地図があり、それぞれを大図・中図・小図と呼んでいます。

(1/36,000 の大図が 214 枚で 1 組)

(1/216,000 の中図が 8 枚で 1 組)

(1/432,000 の小図が 3 枚で 1 組)

地図の作成はまず測量に基づいて大図を作りそれを編集して中図、小図が作られています。

大図は1枚が畳1枚分ほどあり、所領配置や寺院の境内、街道沿いの家並みをはじめ多くの情報が書き込まれています。初期に測量した東日本に比べて何度も訪れて修正が重ねられている西日本の測量のほうが細かくなっています。

## 3. 伊能図作成の方法

測量とは、地表上の各点相互の位置を求め位置、形状を測定して図に表すことです。角度と距離を測ることにより、今いる位置を知ることが出来ます。

伊能忠敬は測量方法として導線法と交会法そして天体観測の3種類を用いました。

導線法は、ある曲り角に目印になる梵天を立て、杖先方位盤で北からの右回りで方位を測り、その梵天間の距離を測定します。

距離の測定には、はじめは麻縄が使われたが水分等による収縮があるため誤差が出やすい為、後に鉄鎖を考案し大いに使用されました。

鉄鎖は一尺（約 30.3 c m）の長さの物を 60 本つないで十間としました。この導線法による測量成果が地図の基本的な線になりました。しかし導線法だけでは誤差が累積してしまうので、誤差を補正する方法として交会法を用いました。

各梵天を立て位置で遠い山の山頂を目標物として半円方位盤で方位を測定します。各梵天で観測した目標物が三カ所であれば三角形、五カ所であれば五角形が出来ます。これが誤差ですので、これを補正して地図を作成するのですが、まだ誤差は吸収しきれません。そこで伊能忠敬が必ず行ったのが天体観測です。

日中の測量を終えて宿に着くと晴れていれば必ず観測地点の緯度を求めます。昼間は太陽の南中高度も観測しました。特定の天体の南中高度を江戸深川の自宅で観測しておき、観測地点で観測した物と高度差を比較して緯度を算出しました。この緯度を確定したことが現在の測量技術と寸分違わぬ日本図を作り上げたと言えます。

ちなみに測量の総日数 3,700 日余りのうち 1,404 日の天体観測を試みたという記録があります。

#### 4. 正確な伊能図

伊能図の正確さの裏付けには、つぎの要件があると思います。

- 1 子午線 1 度の長さが平均 28.2 里（110. 75 k m）という値があります。  
現在の計測数値と比べての **1/1,000** ほどの誤差でしかありません。
- 2 緯度経度の網目の中に実地測量した地形を位置づけました。
- 3 当時としては最新な測量器機を使用し、かつ不備な点があれば改良し測量の精度を高めました。
- 4 地道な作業を正確・正直に成し遂めました。



## **一般講演 3**

**座長: 小島篤博(大阪府立大学)**



## 「伊能大図」に対する GIS データの活用 Utilization of GIS data for “Ino daizu”

菅原 真悟

Shingo Sugawara

TRC-ADEAC 株式会社, 東京都文京区大塚 3-1-1

TRC-ADEAC Inc., 3-1-1, Otsuka, Bukyo-ku, Tokyo

**概要:** 河出書房新社と東京カートグラフィックが共同制作した「デジタル伊能図」は、「伊能大図」214 枚の画像や測量日記などに加えて、関連する GIS データが収められた DVD である。この DVD に含まれる、座標位置情報付き補正画像、測量線、海岸線、宿泊地などの各種 GIS データを GIS ソフトによって解析・再構築し、Web 版「デジタル伊能図」を開発した。この製品はインターネットを介して利用でき、国土地理院が提供する地理院タイルを利用し現代の日本地図と「伊能大図」を重ねたり並べたりして表示する機能を有している。「伊能大図」上に測量線や宿泊地といった地物をプロットすることもできる。ビューアの開発にあたっては、一般的なウェブブラウザ上で閲覧できるようにするため、JavaScript によるオープンソースライブラリである OpenLayers を用いた。一連の開発を通じて得られた知見や課題、課題解決手法などについて報告する。

**Abstract:** Kawade Shobo Shinsha and Tokyo Cart Graphic have produced jointly "Digital Inozu" DVDs. Those contain 214 images of "Ino daizu", a survey diary, and related GIS data. Analyzing and reconstructing various data such as correction images with coordinate position information, survey lines, coastlines, lodging places, and so on, I have developed the web version of "Digital Inozu". This product uses "Ino daizu" and related GIS data and the tiled web map provided online by Geospatial Information Authority of Japan. A current map of Japan and "Ino daizu" can be overlaid or arranged side by side available via the internet. The various features such as survey lines, lodging places, etc. can be plotted on "Ino daizu" too. I have used OpenLayers that is an open source JavaScript library to develop the web version viewer so that those maps and features can be viewed on a general web browser. I report on the knowledge, issues, and problem solving methods obtained through a series of developments.

**キーワード:** GIS, 伊能忠敬, 「デジタル伊能図」, 「伊能大図」, ラスタデータ, ベクトルデータ, GeoTIFF, GeoJSON, 地図タイル, ベクトルタイル, QGIS, tippecanoe, OpenLayers

**Keywords:** GIS, Ino Tadataka, "Digital Inozu", "Ino daizu", Raster data, Vector data, GeoTIFF, GeoJSON, Tiled web map, Vector tile, QGIS, tippecanoe, OpenLayers

### 1. 「デジタル伊能図」とは

江戸時代後期に、わが国で初めて日本全土を実測した伊能忠敬は、「伊能図」と呼ばれる詳細な地図を遺した。河出書房新社と東京カートグラフィックが共同制作した「デジタル伊能図」は、「伊能図」の基本図である「大図」(縮尺 1:36,000、枚数 214 枚)や「江戸府内図」などの画像データ、測量日記の原画像や翻刻テキストなどを収録した DVD であり、スタンダード版とプロフェッショナル版の 2 種類がある[1] [2]。

これらの DVD には、「伊能図」を GIS によって閲覧・分析するためのソフトウェアも収録されている。国

土地院地図と「伊能図」を重ねて表示でき、海岸線、街道、自然地形、地名などが現代地図と比較できる。全国各地の地名や、「伊能図」に記載された宿泊地などから検索でき、宿泊地とリンクした「測量日記」を閲覧することもできる。

さらに、「伊能図」に関する GIS データそのものも収録されており、これらのデータを活用し、閲覧だけでなく独自の解析や分析をすることができる。「伊能大図」214 枚や「江戸府内図」などの原画像はもちろん、現代地図への重ね合わせを意図して幾何補正されたワールドファイル付きのラスタデータ(画像データ)も収録

されている。「伊能大図」の精度は、それ以前に作成された地図とは比較にならないほど高いとはいえ誤差があり、そのままでは現代地図とぴったり重ならない。補正図は、別途調整を行って取得済であった測線を基準に位置合わせがなされ、現代の地図に重なるよう調整がほどこされたものである(参考文献[1]のユーザーズガイド参照)。

補正図のようなラスターデータだけでなく、詳細かつ大量のベクトルデータも収録されている。ベクトルデータは、緯度・経度からなる座標情報と、名称など地物を表すための属性情報からなるデータである。収録されているデータには、点を示すポイント型のデータと線を示すライン型、面を示すポリゴン型のデータがある。内訳は測線(海岸線や街道を測量した線)、海岸線(200年前の海岸線)、宿泊地(測量隊が宿泊した地点)、地名(村名・郡名・国名・寺社名・城郭・自然地名)、図枠(各「大図」の外枠、境界線)である[3]。

「伊能図」は、地理学、地形学、歴史工学、火山学、津波工学などといった分野の研究において利用されている[4]。こうした研究分野はもちろんのこと、これら以外の研究分野にとっても「デジタル伊能図」に収録されている各種デジタルデータは、活用のされかたしだいで、各研究領域に新たな発見をもたらす高いポテンシャルをもった GIS データであると言える。

## 2. Web 版「デジタル伊能図」開発にあたっての課題

ところで、一昨年の 2018 年は伊能忠敬没後 200 年にあたる年であった。これを機に、DVD だけでなくインターネット上でも「デジタル伊能図」を配信し、より多くの人々が「伊能図」を閲覧できるように企画・開発されたものが Web 版「デジタル伊能図」である。現在、ADEAC システム上で利用できる[5][6]。全データの閲覧や全機能の利用が可能な製品版は有料で提供されている。一部の地域(浜松市西側)に限り、重ね機能や測線、地名の表示などを無料で閲覧・利用できる試用版が用意されている[7]。

単に DVD に含まれるソフトやデータを Web サーバやストレージにインストールしたりコピーしたりすれば Web 版として機能するわけではない。まず、DVD に収録されている GIS データをウェブブラウザが読み込める形式のデータに変換する必要がある。次いで、それらのデータをウェブブラウザが表示できるようにするためのビューアを開発する必要もある。DVD 版に収録されている各種データを変換・再構築すること、それらを表示

したりインタラクティブに操作したりできるビューアを開発することが、Web 版を開発するにあたっての主要な課題となった。もちろん他にも地名や測量日記の検索、「伊能図」や「測量日記」の原画像閲覧ビューアなどを含め、Web 版として機能するよう再構築したデータは多岐にわたる。それらの中で、本稿においては、「伊能大図」に対する GIS データの活用事例として、ラスターデータ(画像データ)のタイル化手法、ベクトルデータ(測線、海岸線や地名、宿泊地名などのデータ)のタイル化手法、およびそれらをウェブブラウザ上で閲覧できるようにする OpenLayers を利用したビューア開発の 3 点に関し報告したい。

## 3. ラスターデータのタイル化

ラスターデータをウェブブラウザで利用するために、「伊能大図」の補正図を地図タイル(マップタイル)化した[8]。国土地理院などからオープンデータとして提供されている地図タイルと「伊能大図」を重ねたり並べたりして両者を連動して閲覧できるようにするには、同じ形式の地図タイルを作成する必要がある。



図 1 ラスターデータのタイル化イメージ

変換にあたって、まず、DVD 版に含まれるワールドファイル付き補正画像を GeoTIFF 形式の画像ファイルに変換した。次に、GeoTIFF 画像を入力ファイルとし地図タイル画像ファイルを出力する変換処理を実施した。いずれの変換処理についても QGIS(Windows 版 バージョン 2.18)を利用した[9]。

1 つ目の変換、すなわちワールドファイル付き補正画像を入力ファイルとし、GeoTIFF 形式の画像ファイルを出力ファイルとする変換処理については、QGIS で当該補正画像を開き、別名で保存すればよい。具体的な手順は次のとおりである。QGIS 起動後、「ラストレイヤの追加アイコン」をクリックし、変換元とする画像を開く。空間参照システムの選択を促すダイアログ画面が表示されるので JGD2000 を選択する。選択後、QGIS の画面上に選択した補正画像が表示されるので、画像が正しく読み込まれたか否かを確認でき

る。この時点で、「レイヤパネル」欄に表示されているラスタ画像アイコンを右クリックし、メニューから「名前をつけて保存する...」を選択する。「出力モード」については「画像」を選択し、出力パス(フォルダ名、ファイル名)を指定し、「OK」ボタンをクリックする。

ラスタデータに関する2つめの変換、すなわち、GeoTIFF ファイルを入力ファイルとし地図タイルを出力ファイルとする変換については、QGIS に付属する OSGeo4W というシェルを用いる。具体的な手順としては、同シェルを起動後、変換元とする GeoTIFF ファイルを保存したディレクトリに移動し、gdal2tiles コマンドを実行する。コマンドの実行例は次のとおりである。

```
gdal2tiles -s EPSG:4612 -z 5-  
16 ./inodaizu111.tif ./inodaizu111_tiles
```

印刷の都合で改行されているが実際には改行せずに1行のコマンドとして実行する。

このコマンド例においては、4つのオプションを指定した。-s オプションには、EPSGコードを指定する。ここでは EPSG:4612 とした。-z オプションには、出力したいタイルレベル(ズームレベル)の範囲を指定する。ここでは元画像の画質や出力に要する処理時間を勘案してレベル5から16までを指定した。3つめのオプションには、入力元とする GeoTIFF 画像ファイル名を指定する。4つめのオプションには出力先フォルダ名を指定する。

このように大きく2段階に分けられる一連の変換処理を、「伊能大図」214ファイル分に対して実施した。出力されるタイル画像は PNG 形式のファイルである。1ファイルあたりの容量は数キロバイトから数十キロバイトと小さなファイルとなる一方で、多くのファイルが生成される。ラスタ画像をタイル化することでウェブブラウザが描画に必要な画像データの読み込み量を減じることができ、短時間で地図を描画できる利点がある一方、事前に準備しておくべきファイル数が膨大なものとなる点については注意が必要である。変換前の TIFF 画像に比べて総容量が大きくなるので、利用するハードディスクや SSD などのストレージには十分な空き容量を確保しておいたほうがよい。例えば、元画像として DVD に収録されている「伊能大図」111番「浜松」の TIFF ファイルを変換した場合、変換前のファイルサイズは約 16MB(解像度 300pi、幅 8,349×高さ 4,858 ピクセル)であったが、GeoTIFF ファイル変換後のファイルサイズは約 154MB となり、さらに GeoTIFF ファイルを地図タイル画像に変換後は、総フ

ァイル数 14,227、総フォルダ数 277、総容量は約 490MB となった。

#### 4. ベクトルデータのタイル化

シェープファイルは GIS データフォーマットの一種であり、ベクトルデータを記録するのに適したファイルである[10]。シェープファイルは専用の GIS ソフトウェアを使えばもちろん参照できるが、ウェブブラウザで参照するには、これを GeoJSON など別形式のファイルに変換して利用するのが一般的な手法となる。そこで、ラスタデータの変換においても使用した QGIS を使い、DVD 版「デジタル伊能図」に収録された各シェープファイルを、まずは GeoJSON ファイルに変換した。

シェープファイルを入力ファイルとし、GeoJSON 形式の JSON ファイルを出力ファイルとする変換処理について、具体的な手順は次のとおりである。QGIS 起動後、「ベクタレイヤの追加アイコン」をクリックし、表示されるダイアログ画面上の「ブラウズ」ボタンをクリックして、変換元とするシェープファイルを開く。QGIS の画面上にシェープファイルの内容がプロットされるので、シェープファイルが正しく読み込まれたか否かを確認できる。この時点で、「レイヤパネル」欄に表示されているシェープファイルアイコンを右クリックし、メニューから「名前をつけて保存する...」を選択する。保存設定のためのダイアログ画面が表示されるので、「形式」については「GeoJSON」を選択し、出力ファイル名を所定の入力欄に入力し、「CRS」が EPSG:4612、JGD2000 となっていることを確認し、「OK」ボタンをクリックする。もし出力されたファイルが文字化けするようであれば、事前に GGIS の「設定」メニューから「オプション」メニューの「データソース」を選択し、「シェープファイルのエンコーディング宣言を無視する」のチェックをはずしたうえで再度、変換処理を実行すると、文字化けが解消する可能性がある。

変換後の GeoJSON ファイルについて、それぞれのファイルサイズや要素数は次表のようになった。

表 1 主要ベクトルデータの要素数と GeoJSON ファイル変換後のファイルサイズ

種類	要素数	ファイルサイズ
測線	57,684	230 MB
海岸線	26,796	196 MB
宿泊地	4,471	2 MB
地名	35,521	16 MB
図枠(大図のみ)	214	77 KB

この結果から明らかなように、図枠以外のベクトルデータについては、要素数、ファイルサイズともにウェブブラウザにおいてオンラインの状態では参照するには大きすぎる要素数でありファイルサイズであると言えるだろう。「伊能大図」上にベクトルデータを描画しようとしても、場合によっては、ネットワーク越しに総容量が数百 MB にもなるデータが送受信されることになり、閲覧側の画面はいわゆるフリーズしてしまい、このままでは実用に耐えないことが分かった。

プログラミングに関する格言に、「分割して統治せよ」というものがある。大きな問題は小さく分割して対処すると解決しやすいという意味なのであろうが、この格言をヒントとして、文字どおり、巨大な GeoJSON ファイルを分割する手法を取ればよいことに思い至り、課題解決策として、ベクトルタイル化を試みた[11]。

GeoJSON を入力ファイルとしベクトルタイルを出力ファイルとする変換処理については、mapbox 社が公開する tippecanoe および MBUtil というツールを使用した[12] [13]。どちらも Linux 環境下で動作するコマンドラインツールである。前者のコマンド実行時の出力結果としてベクトルタイルを得ることができ、後者を用いて出力されたファイルを調整する。

tippecanoe コマンドの実行例は次のとおりである。

```
tippecanoe -l sokusen -B9 -al -z16 -Z5 -
o ./sokusen.mbtiles ./sokusen.geojson
```

印刷の都合で改行されているが実際には改行せずに 1 行のコマンドとして実行する。

このコマンド例は、ジオメトリタイプが線の GeoJSON ファイルに対してのものである。この例においては、7 つのオプションを指定した。-l オプションには、ビューアなどに使用する JavaScript が当該ベクトルデータを識別するために必要なレイヤ名を指定する。ここでは sokusen としている。-B オプションには、描画すべきデータが稠密になった際に、適度に表示データを間引く閾値となるタイルレベルを指定する。ここでは 9 を指定している。-al オプションはズームアウト表示時にライン

データを点描するために指定する。ズームアウト時、すなわち、画面上のごく狭い範囲に全情報を使って線を描画すると負荷がかかりすぎ、描画速度が遅くなってしまうのを避けるためである。-z オプションには、描画対象とする範囲の最大ズームレベルを指定する。この例においては 16 を指定している。-Z オプションには、描画対象とする範囲の最小ズームレベルを指定する。この例においては 5 を指定している。同じアルファベットでも小文字の z オプションと大文字の Z オプションでは意味が異なる点に注意が要する。-o オプションには出力先ファイル名を指定する。最後のオプションには入力ファイルとなる GeoJSON ファイル名を指定する。他にもいくつかの有用なオプションがあるので、ジオメトリのタイプやウェブブラウザ側での描画内容、描画の目的、描画速度などの諸条件を勘案しながら、オプションを設定し、出力内容を調整していくとよいだろう。

tippecanoe コマンドによって出力される mbtiles 形式のファイルは、複数のタイルが 1 つのファイルにまとめられたものである。さらにこのファイルを個々のタイルへエクスポートする必要がある。このエクスポートは、mb-util コマンドを用いて、例えば次のように実行する。

```
mb-util --image_format=pbfs sokusen.mbtiles
sokusen_tiles
```

印刷の都合で改行されているが実際には改行せずに 1 行のコマンドとして実行する。--image\_format オプションには出力されるファイル形式を指定する。参考文献[11]によれば、ベクトルタイルは Protocol Buffers 形式のファイル[14]にエンコードされるので、それを示すファイル形式である pbf を指定する。その他、入力ファイルとして mbtiles 形式のファイル名を指定し、最後にタイルの出力先となるルートディレクトリを指定する。

出力された個々のファイルは zip 形式で圧縮されているので、gzip などのコマンドで解凍する。また参考文献[11]によれば、ファイルの拡張子は.mvt が適切であると記載があるので、.mvt という拡張子を解凍後のファイルに付与する。

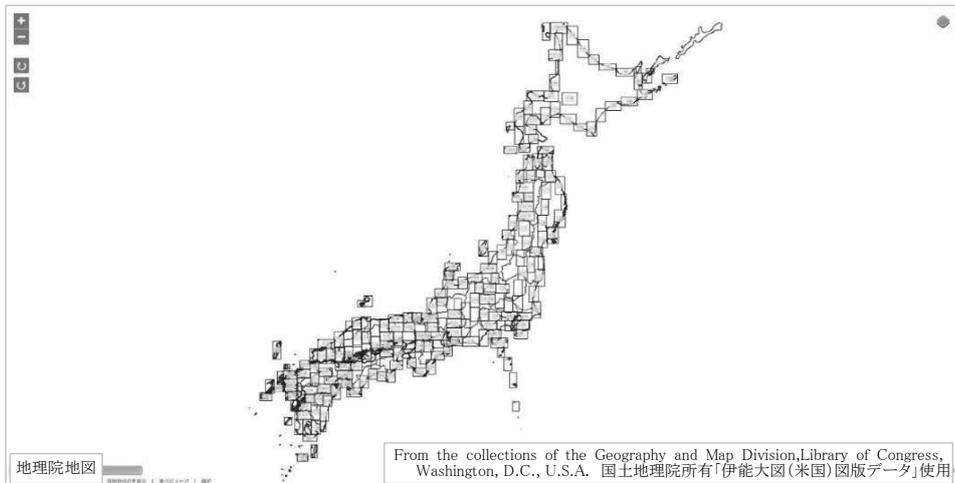


図2 ビューア表示例 地理院タイトル「白地図」に大図 214 枚の図枠をプロット

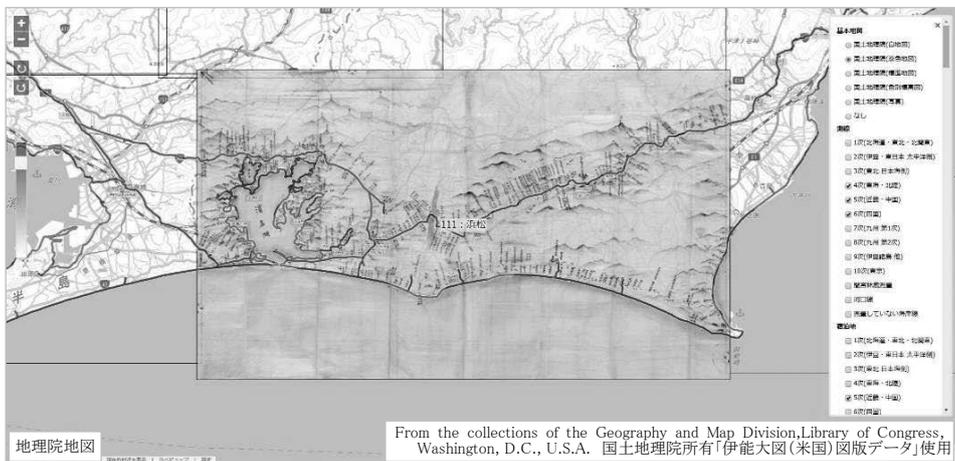


図3 ビューア表示例 地理院タイトル「淡色地図」に大図番号 111 番「浜松」を重ね表示

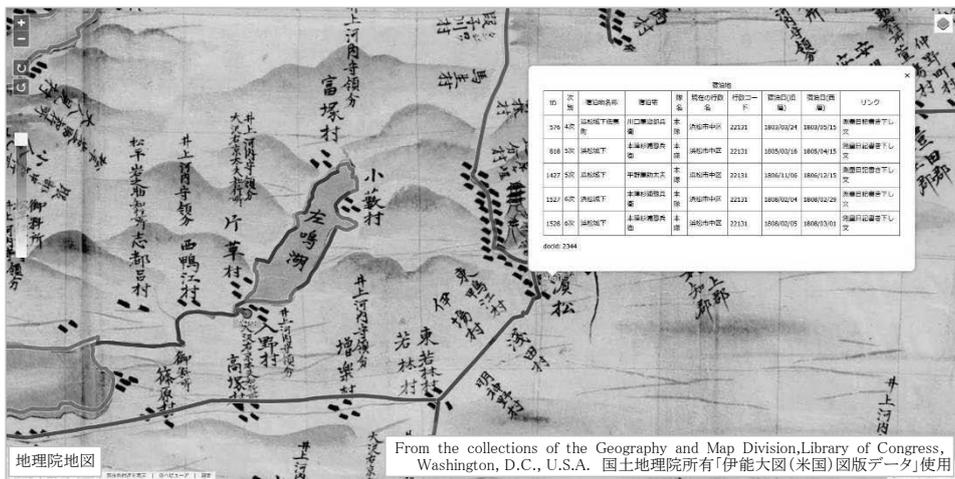


図4 ビューア表示例 宿泊地クリック時のポップアップ情報表示

## 5. OpenLayers

オンライン地図をウェブブラウザで表示するための JavaScript ライブラリとして Leaflet や OpenLayers が知られている[15][16]。Leaflet はシンプルで導入が比較的容易なライブラリとして知られているが、プロットする地物数が多いこと、レイヤを操作するためのコントローラが必要なこと、回転などの複雑な操作が要求されること、ベクトルタイル形式のデータを扱いやすいことなどを考慮し、今回のケースにおいては OpenLayers(バージョン 4.6.5)を利用してビューアを開発した。

OpenLayers を用いて開発したのは、前述したラスターデータやベクトルデータを、国土地理院が配信するいくつかの地図タイルをベースとし、それらに重ねたり並べたりして連動させ、表示することを目的とするビューアである。地図の拡大・縮小・移動・回転、背景とする国土地理院地図の切り替え(白地図・淡色地図・標準地図・色別標高図・写真の 5 種類から選択)、レイヤとして重ねる「伊能大図」214 枚それぞれについて表示・非表示の切り替え、「伊能大図」を現代地図に重ね表示する際の透過率の変更、測線・宿泊地・地名などの表示・非表示選択といった諸機能を、主に実装した。

「伊能大図」ビューアには、以上の基本機能に加えて、画面上に描画された地物をクリックするとベクトルタイルを読み取って属性情報を表示したり、各宿泊地で書かれた「測量日記」を表示するためのリンクを表示したりする機能も実装されている。その他、閲覧するため

に使用している機器が実際に存在する現在位置への移動や、地図の重ね表示・並べ表示切り替え、地名など各種検索結果からの連携表示なども考慮した作りとなっている。

背景地図として描画する国土地理院地図タイルや「伊能大図」214 枚を独自に加工した地図タイルなどのラスターデータは、ol/layer/Tile クラスを使って読み込む。宿泊地や測線などのベクトルタイルは ol/layer/VectorTile クラスを使って読み込む。重ね地図の描画については、これらのクラスを使って読み込んだ各レイヤを、Map クラスのオブジェクトにレイヤ指定すればよい。加えて、透過率を指定するスライダやレイヤを選択するチェックボックス、回転用ボタンなどの別途用意する必要があるコントロールもあわせて Map クラスのオブジェクトに指定する。

図 2～図 4 のような重ね地図を描画する場合、各レイヤを Map クラスのオブジェクトに読み込むだけで、OpenLayers にもともと備わっている機能により、地図の表示位置の移動、拡大縮小、回転などの操作について、基本的には、それぞれの地図の動きや操作が自動的に同期される。しかし、図 5 のように地理院タイルと「伊能大図」タイルを並べて表示する場合は、両地図の動きを同期させる関数を独自に記述しなければならなかった。



図 5 ビューア表示例 地理院タイル「淡色地図」と大図番号 111 番「浜松」の並べ表示

両地図を並べて描画する場合には、双方の地図のどちらか一方において、移動、ズーム、回転といった操作が発生したときに、互いの地図の中心緯度経度、ズームレベル、回転角度について同期をとるようにソースコードを記載すると良い。次の例は地理院地図(地図1とする)の表示上の中心位置が移動した際に「伊能大図」(地図2とする)の表示上の中心位置を同じ位置へ移動させる JavaScript コードの記述例である。

```
1: var v1change = false;
2: var v2change = false;
3:
4: view1.on('change:center', function (evt) {
5:     if (v1change) {
6:         return;
7:     }
8:     if (v2change) {
9:         return;
10:    }
11:    v1change = true;
12:    view2.setCenter(view1.getCenter());
13:    v1change = false;
14: });
```

コード例中の view1 という変数は地図 1 を、view2 という変数は地図 2 を、それぞれ示すものとする。

1 行目に、地図 1 が同期処理中であるか否かを示す変数 v1change を定義している。同様に 2 行目には、地図 2 が同期処理中であるか否かを示す変数 v2change を定義している。

4 行目以降に、地図 1 の中心位置変更イベントに対応する関数を記載している。イベント発生時に処理すべき内容が記載されている関数である。5 行目で、もし地図 1 が同期処理中であれば、以降の処理をキャンセルし、何も処理せずにこの関数を終了する。同様に、8 行目で、もし地図 2 が同期処理中であれば、以降の処理をキャンセルし、何も処理せずにこの関数を終了する。

11 行目で、v1change 変数に true を代入し、地図 1 が同期処理中であることを設定している。12 行目には、この関数の中心となる処理が記載されているのだが、地図 2 側の中心座標に対して、地図 1 と同じ中心座標がセットされている。

13 行目で、v1change 変数に false を代入し、地図 1 側の同期処理が終わったことを設定しているが、これは、地図 1 側での地図の中心位置が変更した際の同期処理を再び受け入れられるようにするためである。

なお、ここに例示したソースコードはあくまでも説明のためのものであり、実際にコピーして利用しても意図したとおりに動作しない可能性がある点に留意されたい。

OpenLayers に関しては、利用できるクラスや API に関するドキュメント、サンプルコードなどが公式サイト上で豊富に公開されているし[17][18]、これら以外の情報源においても有用なコーディング例を見つけ出すことができるとされる。情報源が多いので、どんな API を利用しどのようにコーディングしたらよいかといった点で手がかりを得やすいライブラリである。その一方で OpenLayers の開発が活発に行われていることの裏返しでもあると思われるのだが、開発ペースが予想外に速く、依拠するライブラリのバージョンによってコーディング内容を変えないと、意図した結果を得られない場合もある点には注意が必要である。

## 6. おわりに

「伊能大図」214 枚のラスタデータを地図タイル化するには、QGIS を用いる場合、補正画像(ワールドファイル付き TIFF ファイル)→GeoTIFF ファイル→地図タイルという手順をふめばよいことが分かった。

宿泊地や測線などの地物に関するベクトルデータについては、シェープファイルを GeoJSON ファイルに変換して利用しようとしたところ、ウェブブラウザにとっては一度に送受信されるデータ量が多すぎ、そのままでは描画のために利用できないという問題が生じた。この問題については、ラスタデータ同様に、ベクトルデータもタイル化するのが解決策として有効であった。ベクトルタイルを生成するための変換過程はやや複雑であるが、シェープファイル → GeoJSON ファイル → mbtiles ファイル → pbf ファイル → バイナリベクトルタイルファイルという手順をふめばよいことが分かった。なお、ベクトルタイルについては国土地理院が提供実験を継続しているなど[19]、高度な地図描画にとってますます必要不可欠なものとなっていくことが予想される。

このようにして変換したラスタデータやベクトルデータを、国土地理院からオープンデータとして提供されている地図タイルと連動させて表示したり操作したりするためのビューア開発には、OpenLayers を利用した。OpenLayers に標準で備わっているクラスや API を利用すれば、ほとんどの機能を実装できた。もっとも、一部の機能は独自に実装する必要があり、本稿においては、並べ表示時の同期処理を例示した。

以上のような過程を経ることにより、インターネット上で閲覧できる Web 版「デジタル伊能図」をリリースでき

るという成果が得られた。GIS をめぐる ICT 環境や利用できるデータベース、データセットが、今後、ますます整備されていくにつれ、「伊能大図」に関する様々な活用事例が創発されていくものと予想される。そのような動向に関心を寄せつつ、今後も機会があれば、現在リリースされている Web 版をさらにブラッシュアップしていきたい。

## 参考文献

- [1] 村山祐司, デジタル伊能図 スタンダード版, 河出書房新社, 2015.12
- [2] 村山祐司, デジタル伊能図 プロフェッショナル版, 河出書房新社, 2015.12
- [3] “地図データ定義”.  
<http://www.kawade.co.jp/news/地図データ定義.pdf>, (参照 2020-01-11).
- [4] 河出書房新社編集部 編, モリナガ ヨウ 絵, 伊能図探検, 2018.9, pp. 75-95.
- [5] “ADEAC® : A System of Digitalization and Exhibition for Archive Collections”. <https://trc-adeac.trc.co.jp/>, (参照 2020-01-11).
- [6] “ADEAC とは”. <https://www.trc-adeac.co.jp/about/index.html>, (参照 2020-01-11).
- [7] “河出書房新社・東京カートグラフィック / Web 版デジタル伊能図【お試し版】”.  
<https://trc-adeac.trc.co.jp/WJ11C0/WJJS02U/1391075150>, (参照 2020-01-11).
- [8] “マップ タイル”. <https://www.esri.com/gis-guide/web-gis/map-tile/>, (参照 2020-01-11).
- [9] “QGIS フリーでオープンソースの地理情報システム”. <https://www.qgis.org/ja/site/index.html>, (参照 2020-01-11).
- [10] “シェープファイル”.  
<https://www.esri.com/gis-guide/esri-dataformat/shapefile/>, (参照 2020-01-11).
- [11] “Vector tile specification”.  
<https://docs.mapbox.com/vector-tiles/specification/>, (参照 2020-01-11).
- [12] “tippecanoe”.  
<https://github.com/mapbox/tippecanoe>, (参照 2020-01-11).
- [13] “MBUtil”. <https://github.com/mapbox/mbutil>, (参照 2020-01-11).
- [14] “Protocol Buffers”.  
<https://developers.google.com/protocol-buffers/>, (参照 2020-01-11).
- [15] “Leaflet”. <https://leafletjs.com/>, (参照 2020-01-11).
- [16] “OpenLayers”. <https://openlayers.org/>, (参照 2020-01-11).
- [17] “OpenLayers API”.  
<https://openlayers.org/en/latest/apidoc/>, (参照 2020-01-11).
- [18] “OpenLayers Examples”.  
<https://openlayers.org/en/latest/examples/>, (参照 2020-01-11).

- [19] “ベクトルタイルとその提供実験について”.  
<https://maps.gsi.go.jp/development/vt.html>, (参照 2020-01-11).

本稿を執筆するにあたっては「デジタル伊能図(©東京カートグラフィック 2015、©河出書房新社 2015)」スタンダード版、プロフェッショナル版、両 DVD と、Web 版「デジタル伊能図」を使用いたしました。

## 森鷗外の小説を対象とした文体の継時的な 変化についての計量的な検討

### Quantitative Analysis Of Chronological Changes Of The Writing Style In Ogai Mori's Novels

土山 玄

Gen Tsuchiyama

お茶の水女子大学 文理融合 AI・データサイエンスセンター, 東京都文京区 2-1-1  
Ochanomizu University, 2-1-1, Ohtsuka, Bunkyo-ku, Tokyo

**概要:** 本研究では森鷗外の小説 47 作品を分析対象として、継時的に出現傾向が変化する文体的特徴の抽出を試みた。分析において、用いた特徴量は品詞の比率と助詞、および助動詞の出現率である。これらの特徴量に対して上掲の 47 作品の出版年を目的変数としてランダムフォレストを行い、変数重要度を求めることで継時的に出現傾向が変化する文体的特徴の抽出を行った。分析の結果、1890 年に出版された『うたかたの記』及び『舞姫』、1891 年に出版された『文づかひ』の 3 作品は他の小説と異なる傾向を有していることが明らかになった。これに加えて、形容詞の比率や助動詞の「ない」の出現率などにおいて 1912 年以降の作品ではそれ以前の作品と異なる出現傾向を有する可能性が認められた。

**Abstract:** In this study, we investigate chronological change of writing style of Ogai Mori. He is one of the masterful novelists in Modern Japan, and his literary works are common subjects of the literary research. In this study, we analyse the relative frequency of the words and appearance ratio of the parts of speech using random forests. The results of the analysis indicate that “Utaka no Ki” and “Maihime” which are published in 1890 and “Fumi dukahi” which is published in 1891 are different in writing style from other works. In addition, we reveal that the tendency of word occurrence of auxiliary verb is different before and after 1912.

**キーワード:** 計量文献学、テキストマイニング、機械学習、ランダムフォレスト

**Keywords:** Stylometrics, Text mining, Machine learning, Random forests

#### 1. はじめに

文学作品を対象とし、計量的な手法を用いて文章を分析する研究は計量文献学と称される。計量文献学は、著者の文体に関わる習慣的、形式的特徴を統計的に分析することで著者の識別や推定、文献の成立年代、あるいは成立の順序を推定する学問分野である。このような計量文献学では文体的特徴の出現傾向を調査することで著者の識別や推定を行うことが多く、また数多くの研究成果が報告されている。計量文献学では文体は著者の個性を映しており、文体的特徴は著者間において出現傾向が相違するという考え

に基づいている。文体的特徴とはすなわち文章にあらわれる著者の形式的、あるいは習慣的な表現形式のことである。

また、1 人の作家が多数の作品を残した場合、このような文体的特徴には継時的に出現傾向が変化するものもあることが推測される。つまり、特定の著者の文体的特徴の継時的な変化に注目することで、文体の成長や発展について考察するための透明性の高い資料を提出できると考えられる。

そこで、本研究では日本における文豪として知られる森鷗外の小説を分析対象とし、計量的に分析を行う

ここで継時的に出現傾向が変化すると考えられる文体的特徴を指摘する。本研究では森鷗外の小説について分析を行うにあたって、文体を規定する要素、すなわち文体的特徴であると考えられる、各作品における品詞の比率と単語の出現率を特徴量として多変量解析を行った。特に単語の出現率では助詞及び助動詞を採り上げた。この2品詞を採り上げることについて、助詞及び助動詞は名詞や動詞などと異なり、文中において語彙的意味を担うのではなく文法的機能を担うからである。語彙的意味を担う単語の出現率は小説において描かれるストーリーによって影響されるものと考えられるが、助詞や助動詞などの文法的機能を担う単語の出現率はストーリーによる影響は強くないと考えられる。なお、助詞や助動詞のような語彙的機能を担う単語は機能語と称される。

## 2. 関連研究

日本の文学的文章を対象とし、計量的な手法を用いて著作の執筆順序の推定を目的とした研究では金 (2009) が著名である。金 (2009) では芥川龍之介の著作について分析を行っている。芥川龍之介の文章 309 編を分析対象として採り上げ、統計手法を用いた分析を行った結果、係助詞の「は」及び格助詞の「に」「を」「の」の出現率が継時的に増加し、反対に格助詞の「が」「と」や接続助詞の「て」の出現率が減少していることを明らかにした。

次いで、土山 (2019a) では森鷗外と並び文豪と称される夏目漱石の小説 22 作品を採り上げ、金 (2009) と同様に統計手法を用いて継時的に出現傾向は変化すると考えられる文体的特徴について検討を加えている。土山 (2019a) では夏目漱石の『自然を寫す文章』において「今日では一番言文一致が行はれて居るけれども、句の終りに「である」「のだ」とかいふ言葉があるので言文一致で通って居るけれども、「である」「のだ」を引き抜いたら立派な雅文になるのが澤山ある。」という指摘があることから、文末表現を採り上げ主成分分析を行っている。その結果、文末表現については1908年頃に量的な特徴の変化が認められ、1909年に発表された『それから』以降の作品は文末に助動詞を用いることが増加し、特に文末に助動詞の「た」の使用の増加が顕著であることを指摘している。

また、本研究と同様に森鷗外の小説を対象とした研究も報告されている。森鷗外は1890年から1917年まで作家として活動しているが、表1に示すように『うたかたの記』及び『舞姫』は1890年に、『文づかひ』は1891年に、『そめちがへ』は1897年に発表されている。その後の作品は1909年に発表されていることを考えると森鷗外の初期4作品が発表されてから10年以上の間隔を空けてから他の作品が発表されている。土山 (2019b) では森鷗外の小説 47 作品を対象とし、単語の出現率を特徴量とし計量的な分析を行っている。分析の結果、上掲の『うたかたの記』『舞姫』『文づかひ』の3作品は初期4作品を除く43作品に比べて文語表現、特に文語的な助動詞の出現率が顕著に高く、加えて上掲の3作品ほど顕著ではないが『そめちがへ』も43作品に比べると文語助動詞の出現率が高いことを指摘している。

表1 森鷗外の小説と発表年

タイトル	発表年	タイトル	発表年
うたかたの記	1890	かのように	1912
舞姫	1890	興津弥五右衛門の遺書	1912
文づかひ	1891	鼠坂	1912
そめちがへ	1897	佐橋甚五郎	1913
キタ・セクスアリス	1909	護持院原の敵討	1913
半日	1909	阿部一族	1913
鶏	1909	堺事件	1914
あそび	1910	大塩平八郎	1914
普請中	1910	安井夫人	1914
木精	1910	栗山大膳	1914
杯	1910	じいさんばあさん	1915
沈黙の塔	1910	二人の友	1915
牛鍋	1910	余興	1915
独身	1910	山椒大夫	1915
花子	1910	最後の一句	1915
里芋の芽と不動の目	1910	津下四郎左衛門	1915
青年	1910	魚玄機	1915
食堂	1910	伊沢蘭軒	1916
カズイステカ	1911	壽阿彌の手紙	1916
妄想	1911	寒山拾得	1916
心中	1911	相原品	1916
百物語	1911	渋江抽斎	1916
雁	1911	高瀬舟	1916
		細木香以	1917

たの記』及び『舞姫』は1890年に、『文づかひ』は1891年に、『そめちがへ』は1897年に発表されている。その後の作品は1909年に発表されていることを考えると森鷗外の初期4作品が発表されてから10年以上の間隔を空けてから他の作品が発表されている。土山 (2019b) では森鷗外の小説 47 作品を対象とし、単語の出現率を特徴量とし計量的な分析を行っている。分析の結果、上掲の『うたかたの記』『舞姫』『文づかひ』の3作品は初期4作品を除く43作品に比べて文語表現、特に文語的な助動詞の出現率が顕著に高く、加えて上掲の3作品ほど顕著ではないが『そめちがへ』も43作品に比べると文語助動詞の出現率が高いことを指摘している。

## 3. データ

本研究に用いた森鷗外の小説は表1に示した1890年から1917年までに発表された47作品である。なお、これら47作品は上掲の土山 (2019b) と同じである。

また、これらの小説のテキストデータは web サイトの青空文庫 (<http://www.aozora.gr.jp/>) から入手した。

次に、それらのテキストデータに対し、形態素解析によって単語に品詞のタグ付けを行った。形態素解析は MeCab ver. 0.996 を、形態素解析の際に用いる辞書は UniDic ver. 2.0.1 を用いた。

このような処理によって作成されたテキストデータを対象に統計的な分析を行った。分析に際して、先にふれたように品詞の比率と助詞及び助動詞の出現率を特徴量として用いた。品詞の比率は作品別に各品詞の頻度を集計し、各作品の延べ語数に対する割合を求めた。次に、単語の出現率は品詞の比率と同様に作品別に各単語の頻度を集計し、各作品における品詞別の総度数に対する割合を求めた。

## 4. 分析

### 4.1 分析手法

本研究では分析において、主にランダムフォレストを用いた。ランダムフォレストを用いることで森鷗外の小説 47 作品において出現傾向が継時的に変化する文体的特徴を抽出した。ランダムフォレストとは機械学習の手法の 1 つであり、決定木あるいは回帰木のアンサンブル学習とも言える分析手法である。ランダムフォレストでは、まず分析対象の個体数の 2/3 にあたるブートストラップサンプルを抽出し、そのブートストラップサンプルを対象とし未剪定の決定木あるいは回帰木を生成する。また、未剪定の木を生成する際に、すべての変数を用いず、一般的に変数の数の平方根にあたる数の変数を用いる。ランダムフォレストはこのような未剪定の決定木あるいは回帰木を大量に生成し、分析結果を統合することで最終的な結果を得る。従って、ランダムフォレストを繰り返すと、同一の結果が得られることはおおよさない。本研究では表 1 に示した出版年を目的変数としてランダムフォレストを行った。

また、ランダムフォレストでは分析を行う上で変数の重要度を推定する。本研究では森鷗外の小説の出版年を目的変数としており、出版年の推定における変数重要度が求められる。従っ

て、この変数重要度が高い変数が森鷗外以外の 47 作品の小説において出現傾向が継時的に変化している文体的特徴であると考えられる。

### 4.2 分析結果

本研究ではまず品詞の比率について分析を行った。分析では先にふれた Mecab 及び Unidic を用いた形態素解析においてタグ付けされた品詞のタグを用いた。具体的には名詞、代名詞、動詞、形容詞、形

状詞、副詞、連体詞、接続詞、感動詞、助詞、助動詞、接頭辞、接尾辞、補助記号、記号の 15 のタグである。なお、形状詞は形容動詞の名詞語根に相当する。補

表 2 47 作品を対象とした品詞の変数重要度

品詞	重要度
感動詞	656.203
連体詞	223.859
助動詞	211.591
記号	147.152
補助記号	97.573
形容詞	71.985
接続詞	65.278
形状詞	62.134
助詞	44.009
代名詞	42.878
名詞	32.190
副詞	30.031
接尾辞	29.360
動詞	27.646
接頭辞	11.721

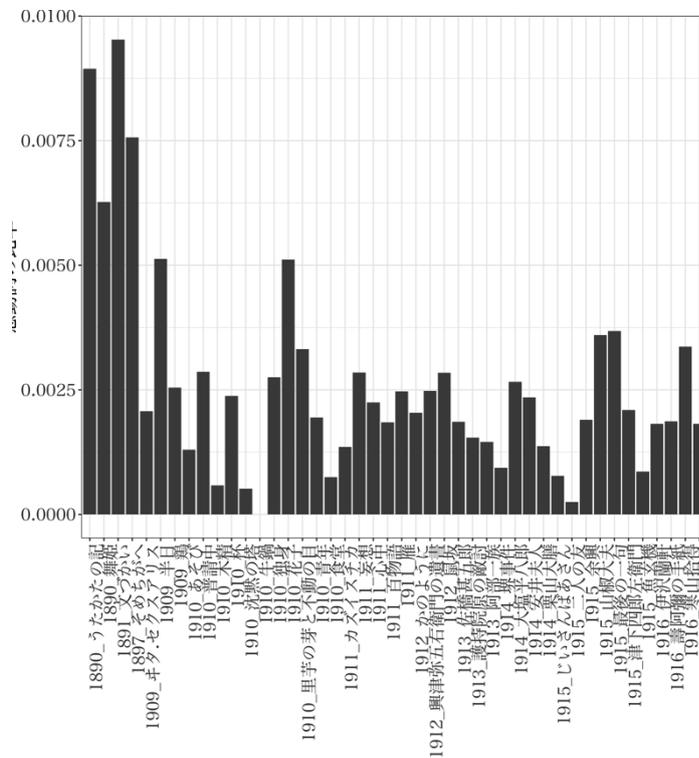


図 1 各作品における感動詞の比率

助記号は句読点やかぎ括弧などが含まれ、記号は文法的な機能を担わない記号が該当する。

これらの 15 のタグを説明変数とし、47 作品の出版年を目的変数としランダムフォレストを行った。その結果、表 2 に示すように推定された変数重要度は感動詞、連体詞、助動詞が高い。よって、森鷗外の小説 47 作品において、これら 3 品詞の出現傾向が継時的に変化している可能性が考えられる。そこで、変数重要度が最も大きかった感動詞の各作品における比率を可視化すると、図 1 に示すように初期の作品において感動詞の比率が高く、それ以降の作品に継時的な変化は認められないと考えられる。

次いで、15 の品詞タグを用いて、相関行列を用いた主成分分析を行った。図 2 は主成分分析によって求められた主成分得点の散布図である。横軸が第 1 主成分を、縦軸が第 2 主成分を意味している。なお、第 1 主成分の寄与率は 41.8%、第 2 主成分の寄与率は 13.0%であり、第 2 主成分までの累積寄与率は 54.9%である。図 2 において、初期 3 作品である『うたかたの記』『舞姫』『文づかひ』の第 2 主成分の主成分得点が小さく、これら 3 作品は類似した傾向を有していると考えられる。第 2 主成分の主成分負荷量は表 3 に示す通りであり、初期 3 作品は感動詞及び連体詞の比率が高く、助動詞の比率が小さい作品群であると解釈される。これは表 2 に示したランダムフォレストの結果と合致する。

よって、初期 3 作品を除き 44 作品を対象として改めてランダムフォレストを行った。表 4 はランダムフォレストの結果として得られた変数重要度であり、形容詞の重要度が最大となった。これら 44 作品の形容詞の比率は図 3 に示す通りである。形容詞の比率は単調な変化を示していないが、1912 年以降の作品では顕著に形容詞の比率が認められる。従って、ここに森鷗外の小説における 1 つの文体的特徴の継時的な変化が明らかになったと言える。

次に、小説 47 作品を対象とした単語の出現率を特徴量としてランダムフォレストを行った。先に述べたように本研究では助詞と助動詞を採り上げ、分析を行った。まず助詞の出現率に対してランダムフォレストを行った。分析によって求められた変数重要度は表 5 に示す通

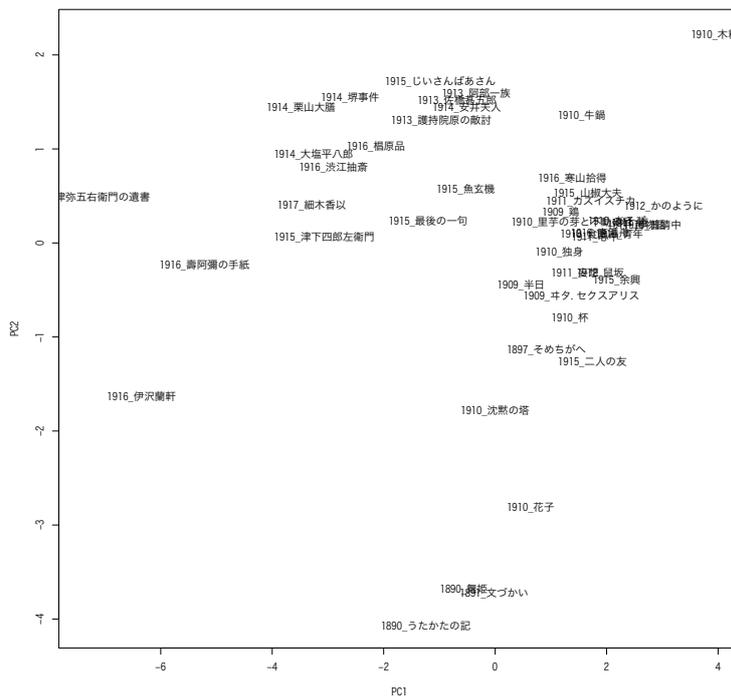


図 2 品詞の比率の主成分分析の結果

りである。「のみ」の変数重要度が最大となり、「し」及び「ど」などの変数重要度も高い。そこで、「のみ」の各作品における出現率を可視化すると、図 4 に示すように初期 4 作品における出現率が顕

表 3 主成分負荷量

	PC1	PC2
名詞	-0.384	0.078
助詞	0.356	0.072
動詞	0.349	0.096
補助記号	0.067	-0.298
助動詞	0.239	0.228
接尾辞	-0.356	-0.033
副詞	0.310	-0.027
代名詞	0.164	-0.209
接頭辞	-0.235	-0.020
形容詞	0.327	-0.042
連体詞	0.072	-0.477
形状詞	0.346	0.015
接続詞	-0.062	0.106
記号	-0.032	-0.545
感動詞	0.012	-0.506

著に高く、その他の作品ではおよそ出現しない。そこで、品詞に対する分析と同様に、初期 3 作品を分析対象から除外し、改めてランダムフォレストを行った。その結果、「に」「か」「も」と言った助詞の変数重要度が高く推定された。図 5 は各作品における「も」の出現率であり、1912 年より出現率が減少傾向にあると考えられる。

表4 44作品を対象とした品詞の変数重要度

品詞	重要度
形容詞	106.736
感動詞	93.422
補助記号	92.383
形状詞	31.325
助詞	27.502
接続詞	25.297
代名詞	20.821
連体詞	19.389
記号	14.778
接尾辞	12.550
名詞	8.121
助動詞	8.056
副詞	5.664
動詞	4.352
接頭辞	4.226

表5 47作品を対象とした助詞の変数重要度

	重要度
のみ.助詞	150.178
し.助詞	144.760
ど.助詞	140.729
ば.助詞	136.973
など.助詞	136.318
こそ.助詞	121.418
とて.助詞	121.092
しき.助詞	117.250
にて.助詞	72.315
より.助詞	70.872
に.助詞	66.114
きに.助詞	57.908
なり.助詞	55.214
から.助詞	48.221
で.助詞	38.027

表6 47作品を対象とした助動詞の変数重要度

	重要度
ず.助動詞	333.909
たる.助動詞	328.051
た.助動詞	255.677
なり.断定.助動詞	194.917
たり.断定.助動詞	165.736
ない.助動詞	105.531
だ.助動詞	71.238
べし.助動詞	52.153
ごとし.助動詞	50.207
や.助動詞	49.674
り.助動詞	32.982
てる.助動詞	20.603
しめる.助動詞	14.139
たい.助動詞	13.628
まじ.助動詞	11.138

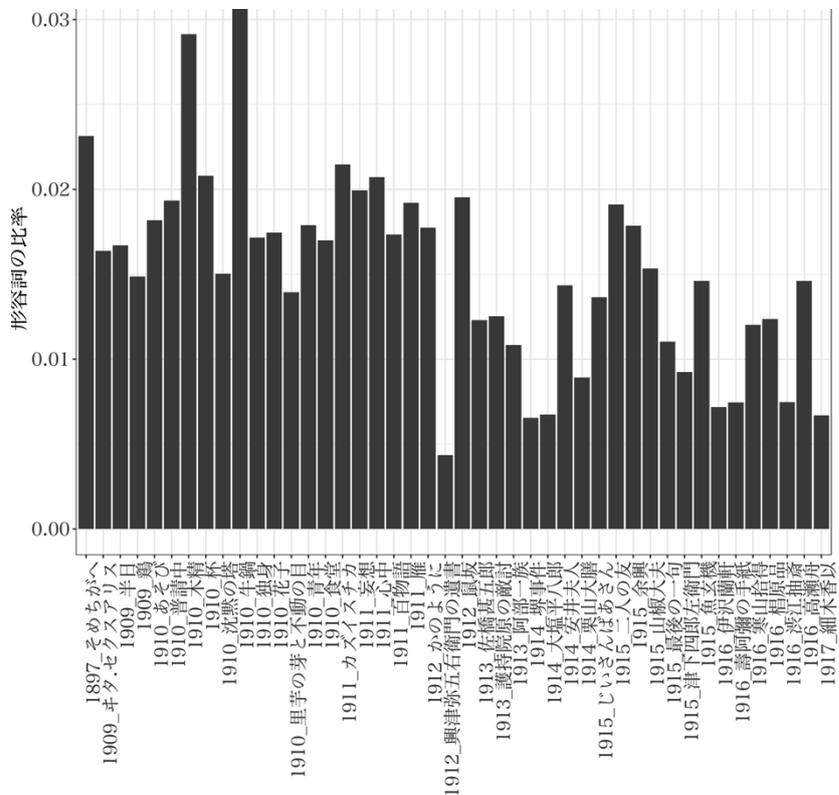


図3 各作品における形容詞の比率



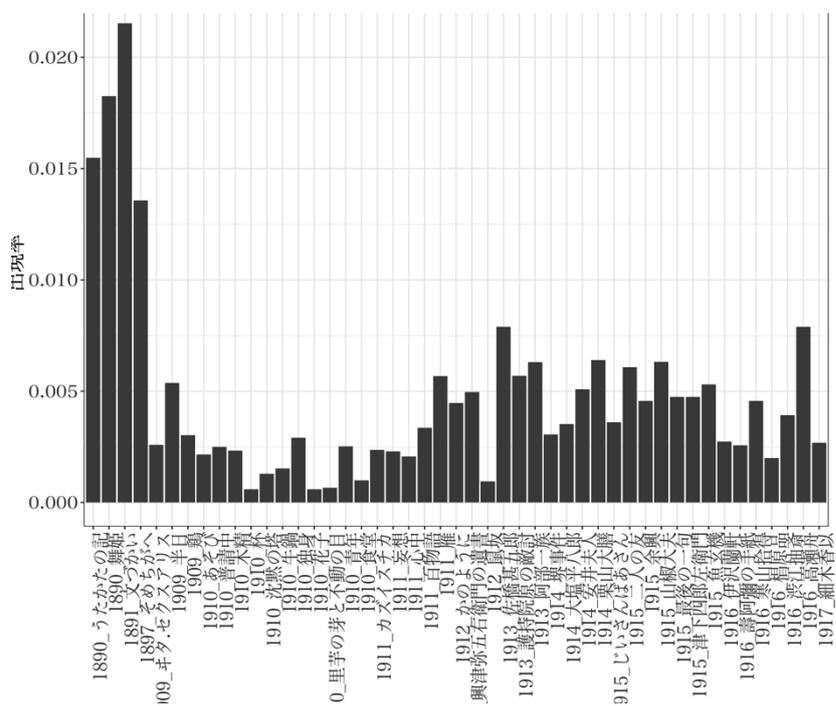


図6 各作品における助動詞「ず」の出現率

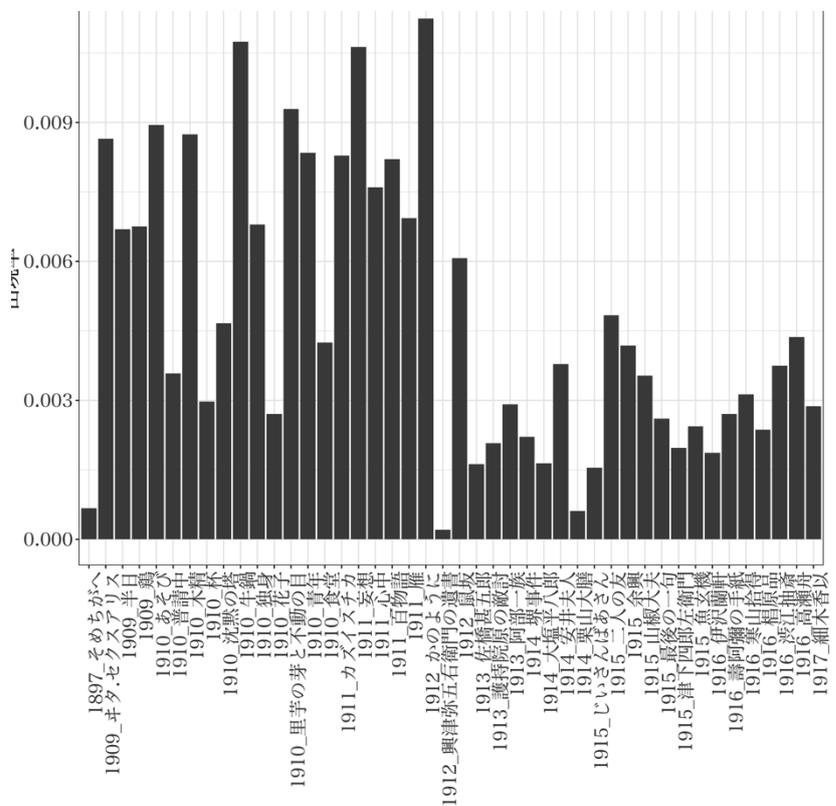


図7 各作品における助動詞「ない」の出現率

最後に、助動詞の出現率についてランダムフォレストを行った。分析の結果として求められた変数重要度は表 6 に示す通りである。表 6 より「ず」「たる」と言った助動詞の重要度が高く推定された。しかし、図 6 において示した「ず」の出現率のグラフのように、これまでの分析と同様に初期の作品における出現率が偏って高いため、このような結果となったと考えられる。そこで、初期 3 作品を分析対象から除外し、改めてランダムフォレストを行った。その結果、表 7 に示すように「ず」「たい」「ない」と言った単語の重要度が高く推定された。図 7 は「ない」の各作品における出現率を可視化したグラフである。図 7 においても 1912 年以降の作品における出現率の低下が認められると考えられる。

## 5. 考察

本研究では、森鷗外の小説 47 作品を対象に、機械学習の手法の 1 つであるランダムフォレストを用い、出現傾向が継時的に変化する文体的特徴の抽出を試みた。その結果、品詞の比率の分析、助詞及び助動詞の出現率に対する分析において、出現傾向が変化する文体的特徴が明らかになった。従って、継時的に出現傾向が変化する文体的特徴の抽出を目的とするとき、ランダムフォレストは有効な分析手法の 1 つであると考えられる。

また、本研究における品詞に対する分析及び単語の出現率に対する分析によって、1890 年に出版された『うたかたの記』及び『舞姫』、1891 年に出版された『文づかひ』の 3 作品は他の小説と異なる傾向を有していることが明らかになった。これに加えて、形容詞の比率や助動詞の「ない」の出現率などにおいて 1912 年以降の作品ではそれ以前の作品と異なる出現傾向を有する可能性が認められた。

## 参考文献

- [1] 漱石全集第 34 巻. 岩波書店, 1957.
- [2] 金明哲. 文章の執筆時期の推定—芥川龍之介の作品を例として—. 行動計量学, 2009, Vol. 36, No. 2, pp. 89-103.
- [3] 土山玄. 夏目漱石の小説における文語表現について. じんもんこん 2018 論文集, 2018, Vol. 2018, pp. 269-276.
- [4] 土山玄. 文末表現の計量分析に基づく夏目漱石の小説の分類. 研究報告人文科学とコンピュータ, 2019a, 2019-CH-120, Vol. 6, pp. 1-4.
- [5] 土山玄. 森鷗外の文体的特徴の変化に関する計量的な考察. 人文・自然研究, 2019b, Vol. 13, pp. 107-115.
- [6] 工藤彰; 村井源; 往住彰文. 計量分析による村上春樹長篇の関係性と歴史の変遷. 情報知識学会誌, 2011, Vol. 21, No. 1, pp. 18-36.

## ADEAC の画像データを利用したくずし字認識 AI の開発と 組み込みシステムへの実装

### Development of Embedded AI System for Recognition of *Kuzushiji* by Using the Image Data on ADEAC®

早坂 太一<sup>1</sup> 竹内 正広<sup>1</sup> 大野 亙<sup>1</sup>

加藤 弓枝<sup>2</sup> 山本 和明<sup>3</sup> 石間 衛<sup>4</sup> 石川 徹也<sup>4</sup>

Masahiro Takeuchi<sup>1</sup>, Taichi Hayasaka<sup>1</sup>, Wataru Ohno<sup>1</sup>,

Yumie Kato<sup>2</sup>, Kazuaki Yamamoto<sup>3</sup>, Mamoru Ishima<sup>4</sup>, and Tetsuya Ishikawa<sup>4</sup>

<sup>1</sup>豊田工業高等専門学校, 愛知県豊田市栄生町 2-1

<sup>1</sup>National Institute of Technology, Toyota College, 2-1 Eisei, Toyota, Aichi

<sup>2</sup>鶴見大学, 神奈川県横浜市鶴見区鶴見 2-1-3

<sup>2</sup>Tsurumi University, 2-1-3 Tsurumi, Tsurumi-ku, Yokohama, Kanagawa

<sup>3</sup>国文学研究資料館, 東京都立川市緑町 10-3

<sup>3</sup>National Institute of Japanese Literature, 10-3 Midori-cho, Tachikawa, Tokyo

<sup>4</sup>TRC – ADEAC株式会社, 東京都文京区大塚 3-1-1

<sup>4</sup>TRC-ADEAC, Inc., 3-1-1 Ohtsuka, Bunkyo-ku, Tokyo

**概要:** 古典籍におけるくずし字翻刻に関する研究では深層学習を利用したアプローチが盛んである。本研究では、ROIS-DS 人文学オープンデータ共同研究センターが公開している 100 万字を超えるくずし字データセットにデジタルアーカイブシステム ADEAC 内の古典籍画像から抽出した字形データ約 2 万 6 千字を加えて、16 ビット Unicode にくずし字を分類する畳み込みニューラルネットワークの学習を行った。その結果、版本のテストデータではそれほど影響が見られなかった認識率について、写本のテストデータに対しては明らかな向上が見られた。さらに、そのモデルをシングルボードコンピュータ Raspberry Pi に実装することで、複数のくずし字を一括して自動検出し、認識を行うことのできる組み込みシステムを開発した。インターネットへの接続を必要としないため、小中学校での教育や古民家での調査などの場面で手軽に利用でき、くずし字翻刻の支援ツールとして活躍することが期待できる。

**Abstract:** There are many approaches using deep learning in research on the interpretation of *kuzushiji* characters in Japanese ancient documents. In this study, we trained a convolutional neural network that classifies *kuzushiji* into 16-bit Unicode characters. In addition to the over 1 million characters published by Center for Open Data in the Humanities, we extracted approximately 26,000 *kuzushiji* characters as learning data from the Japanese ancient document images on ADEAC®, which is the system of digitalization and exhibition for archive collections. The recognition rate was clearly improved for the test data in handwritten documents, which was not significantly affected for the test data in printed documents. Furthermore, we developed the embedded system that can automatically detect and recognize multiple *kuzushiji* by implementing the deep learning model on the single-board computer Raspberry Pi. Since the system does not require the internet connection, it can be expected to play an active role as a support tool for interpreting *kuzushiji* in the situations such as education in elementary and junior high schools, surveys in old houses, and so on.

**キーワード:** 文字認識, くずし字, ADEAC, 深層学習, 組み込みシステム

**Keywords:** character recognition, *kuzushiji*, ADEAC, deep learning, embedded system

## 1. まえがき

国文学研究資料館により平成 26 年度より開始された「日本語の歴史的典籍の国際共同研究ネットワーク構築計画」[1]では、研究基盤整備として約 30 万点の歴史的典籍を画像データ化し、既存の書誌情報データと統合させたデータベースの構築を行っている。あらゆる分野の書籍が含まれる膨大な画像データを有効活用できれば、例えば津波や噴火等の天変地異の歴史を教訓とした防災研究のように、人文科学のみならず自然科学系分野を融合させた研究の展開も期待される。しかしながら多くの研究者にとっては、それらに書かれている文字が「くずし字」であることが障壁となる。

現行のくずし字翻刻に関する研究の中で、コンピュータ技術によるくずし字自動翻刻に関する研究は最も先行研究の蓄積があり(例えば[2][3])、進捗度の大きい分野であると考えられる。著者らの研究グループは深層学習を用いたアプローチの先駆け[3]としてくずし字の自動翻刻の試みを行っている。本稿では、ROIS-DS 人文学オープンデータ共同研究センターが公開しているくずし字データセット[4]に、デジタルアーカイブシステム ADEAC [5]の古典籍画像から抽出した字形データを加えた学習データを用いて、くずし字を分類する畳み込みニューラルネットワークの学習を行い、その結果から、版本および写本に対する学習データの相違による認識精度を比較する。さらに、小型で比較的安価なシングルボードコンピュータ Raspberry Pi にその学習モデルを組み込んだくずし字の自動検出および認識を行うシステムについて報告する。

## 2. 人工知能によるくずし字翻刻

コンピュータ技術によるくずし字自動翻刻に関する研究に分類される本研究において用いる深層学習 (deep learning) は、第3次AIブームの中心的技術として様々な分野で導入が進んでいる。その原理は、ヒト脳内における多数の神経細胞による情報のやりとりを数式によりモデル化した階層型ニューラルネットワーク (多層パーセプトロン) が基になっている。深層学習によりモデルを構築するには、GPGPU (General-Purpose computing on Graphics Processing Units) といった計算機技術の導入を必要とするが、一度モデルを構築しさえすれば階層型ニューラルネットワークと同様に自動翻刻に要する時間はごく僅かである。また、学習に用いる文字画像を多数用意する必要はあるが、学習後のモデルにはそれぞれの古典籍やそれらが書かれた時代で異なる可能性のあるくずし字の特徴が反映されているため、翻刻の際に膨大なデータベースを用意する必要はない。つまり、人工知能技術の導入によっ

て、クラウドコンピューティングや第5世代移動通信システム(5G)に頼ることのない、一般的に普及している携帯情報端末やパーソナルコンピュータでも動作する小規模なアプリケーション・ソフトウェアとして「いつでも/どこでも/誰でも自動翻刻」を実現することが可能になると考えられる。

## 3. 深層学習によるくずし字認識モデルの学習

本研究では深層学習で得られたくずし字認識モデルを用いる。このモデルは4層の畳み込み層と3層の全結合層から構成される畳み込みニューラルネットワーク (CNN; iConvolutional Neural Network) である。くずし字を一文字ずつ 64×63 ピクセルの大きさにリサイズし、ネガ・ポジを反転した JPEG 形式のグレイスケール画像を入力とし、16 ビット Unicode 65,535 クラスに分類する学習を行った。CNN に漢字認識を行わせた先行研究[13]を参考にし、本研究で用いたネットワーク構造を表 1 に示す。

表 1 学習に用いたネットワーク構造 (<sup>1</sup>+Max Pooling 層, <sup>2</sup>+ReLU 層, <sup>3</sup>+Dropout 層)

層の種類	ユニット数
畳み込み層 <sup>1,2</sup>	64
畳み込み層 <sup>1,2</sup>	128
畳み込み層 <sup>1,2</sup>	256
畳み込み層 <sup>1,2</sup>	512
全結合層 <sup>2,3</sup>	1,024
全結合層 <sup>2</sup>	1,024
全結合層(出力)	65,535

一連の数値計算は画像認識用として代表的な深層学習用ライブラリである Caffe [6]を用いて行われた。計算機環境として、OS は Ubuntu 14.04 LTS, CPU は Intel Core i7 6900K 8core / 16thread 3.2GHz, 主メモリは 128GB (16GB×8) DDR4-2133, GPU は nVidia GeForce 1080Ti 11GB×2 を搭載したワークステーション GDEP Deep Learning Box を利用した。

学習には、ROIS-DS 人文学オープンデータ共同研究センター (CODH; Center for Open Data in the Humanities) が公開しているくずし字データセット[4] (以下、CODH データと略す)、および ADEAC [5]における「常総市デジタルミュージアム」「宮代町デジタル郷土資料」「京都女子大『山城国淀藩上月家文書』」から提供された 4,335 種類、1,117,703 文字のくずし字画像データを用いた。学習データの詳細を表 2 に示す。また、表 3 に学習には用いていないテストデータに対する認識率、すなわち確信度の高い文字があらかじめわかっている翻刻結果と同じであった割合を示

表2 学習データの詳細

	ダウンロード		手作業で抽出			小計	種類数	種類数／文字
	CODHデータ[4]	和翰名苑[7]	ADEAC[5]	CODHデータ[8]	五體字類[9]			
変体仮名 (割合)	714,568 (96.82%)	3,265 (0.44%)	9,209 (1.25%)	9,511 (1.29%)	1,473 (0.20%)	738,026	77	9,585
漢字 他 (割合)	360,550 (94.96%)	—	16,920 (4.46%)	2,207 (0.58%)	—	379,677	4,258	89
小計 (割合)	1,075,118 (96.19%)	3,265 (0.29%)	26,129 (2.34%)	11,718 (1.05%)	1,473 (0.13%)	1,117,703	4,335	258
種類数	4,299	48	1,638	417	48			

表3 テストデータに対する認識結果

(上段: 確信度最上位, 下段: 確信度 10%以上のすべての文字に対する認識率)

	(版本)CODH 源氏物語[10]	認識率		(写本)ADEAC 御着城御当日 御規式帳[11]	認識率	
		上記すべて 学習に使用	CODHデー タ[4]使用		上記すべて 学習に使用	CODHデー タ[4]使用
変体仮名	9,852	98.15%	98.08%	2	50.00%	100.00%
		99.39%	99.37%		100.00%	100.00%
漢字 他	1,280	87.58%	87.66%	964	74.07%	66.22%
		93.75%	92.97%		81.95%	73.34%
小計	11,132	96.94%	96.88%	966	73.91%	65.94%
		98.74%	98.63%		81.78%	73.19%

す. 本研究では, CODH データのみを学習に用いた場合と, ADEAC における古典籍画像データから手作業で抽出した字形データをそれらに加えた場合とで, 江戸時代の版本である『源氏物語』[10], および写本である『御着 城御当日御規式帳』[11]をテストデータとしたときの認識率を比較した.

CODH オープンデータが 100 万字を超えるデータ数なのに対して, ADEAC からの字形データは 2 万 6 千字あまりと, ほとんど学習に影響がないように予想されるが, 表 3 より, 版本のテストデータではそれほど影響が見られなかった認識率について, 写本のテストデータに対しては明らかな向上が見られていることがわかる. これは ADEAC のデータが全国各地の写本を主とするものであることによるものと考えられる. 割合としては少なくとも, 学習する字形データの時代や書き手, 種類などのバリエーションを増やすことは, こうした深層学習によるくずし字翻刻アプローチには必要なことであると考える.

学習したモデルを利用して, 古典籍の画像データを読み込み, マウスや指で選択された1文字分のくずし字を翻刻する WWW アプリケーションが 2016 年より公

開されている[3] (<http://vpac.toyota-ct.ac.jp/kuzushiji/>). ブラウザ画面の例を図 1 に示す.



図1 CNNによるくずし字認識用 WWWアプリケーションのスクリーンショット例

このアプリケーションでは、画像ファイルの形式にこだわることなくスマートフォン等で撮影した画像で手軽にくずし字を調べることができる。WWW サーバとして Apple Mac Mini を用い、GPU ではなく CPU による認識を行っている。表示についてはクライアント側の計算機環境に依存するが、サーバ側で1文字あたりの分類にかかる時間は約 0.4 秒である。

## 4. Raspberry Pi を用いたくずし字自動検出・認識システム

### 4.1 システムの概要

現状では、多人数が利用できる Wi-Fi 等のインターネット環境が整備されている小中学校やコミュニティ施設はそう多くはないため、古典籍に関する生涯学習等の講座を開催したい場合に、そうした環境を事業提供者が用意することは相応の時間と費用がかかる。本研究では、インターネット環境を必要としないくずし字の自動抽出および認識を行う組み込みシステムを開発した。図 2 にシステムの概要図を示す。

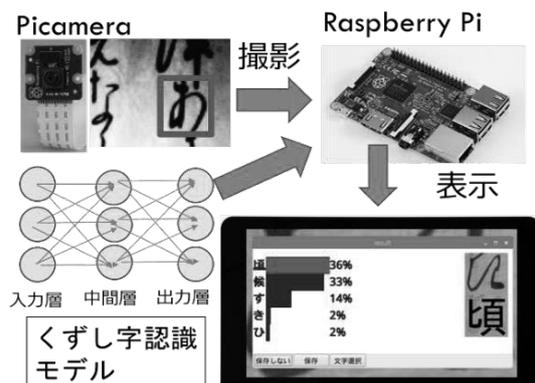


図 2 開発システムの概要図

図 2 に示すシステムでは、ハードウェアとして小型で比較的安価な教育用シングルボードコンピュータ Raspberry Pi Model 4B を用いた。ディスプレイの背面に設置した Raspberry Pi 用カメラモジュールから対象の古典籍画像を撮影できる。また、Raspberry Pi に挿入された SD カードに保存された画像の読み込みも可能である。

§ 3 で述べた CNN モデルをシステムに搭載しているが、これはくずし字1文字に対して認識を行うものであるため、Python3 および OpenCV3 を用いて認識するくずし字の選択や領域検出のための画像処理などを行う。認識は機械学習用フレームワーク TensorFlow Lite [12] を用いて行われ、確信度の高い順に五つの認識結果の候補がタッチスクリーンに表示される。

### 4.2 くずし字領域の検出

古典籍画像に対する前処理には OpenCV3 を利用した。1文字の選択では、タッチスクリーンに触れてから離すまでのx座標およびy座標を取得した後、その領域をリサイズし、2値化を行った後に、CNN モデルに入力している。しかしながら、1文字ずつ手作業で抽出する方法では手間がかかるため、古典籍画像から自動でくずし字領域を検出する機能を実装した。

くずし字領域の自動検出にはモルフォロジ処理を利用した。入力画像に対して、図 3(a)に示すように縦方向に対して膨張処理を行い、上下の字を繋げることで、1行ごとの領域検出を行う。次に、図 3(b)に示すように横方向に膨張処理を行い、それぞれの行から1文字ずつの領域を検出する。図 3(c)に検出されたくずし字領域の例を示す。

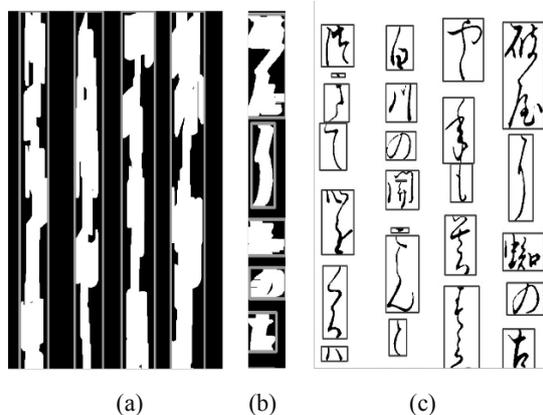


図 3 文字領域の検出過程と結果例

また図 4 に示すように、極端に小さい領域は除外すると共に、領域候補ごとに CNN モデルで一度認識を行い、確信度の高いものを採用することで、領域の統合または分割を行っている。100 文字あまりの古典籍画像に対して約 5 秒でくずし字領域の検出が可能である。

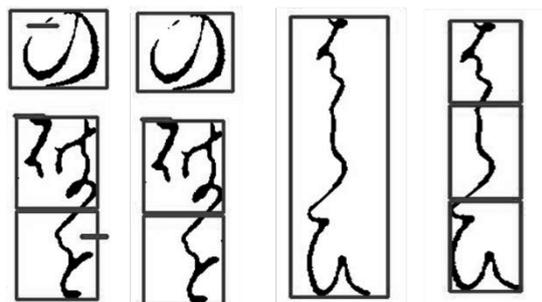


図 4 検出した文字領域の分割・除外・統合の例

### 4.3 モデルの量子化

本研究では、Raspberry Pi による推論用の深層学習用ライブラリとしてモバイル端末や組み込み機器向けの TensorFlow Lite [12]を使用した。TensorFlow Lite では、学習済みモデルの重み等を浮動小数点形式から固定小数点形式に量子化し、モデルサイズを小さくすることで、処理能力がやや劣る機器での演算および読み込みにかかる時間を短縮することができる。本研究では、入力のみ 32 ビットの浮動小数点形式とし、他のパラメータは 8 ビットの固定小数点形式に量子化した。その結果、認識精度は量子化前と同程度のままモデルサイズは約 1/4 に軽量化され、1文字あたりの認識にかかる時間は約 0.08 秒とすることができた。

### 4.4 操作例

本システムでは、GUI アプリケーションから、カメラによる古典籍画像撮影、ファイル読み込み、文字領域抽出、手動での切り取り、自動認識のそれぞれの処理を行うことができる。

カメラモジュールは3秒間のプレビュー後に古典籍画像を撮影するよう設定した。SDカードやUSBフラッシュメモリに保存済みの画像ファイルの読み込みを行うことも可能である。文字認識後は図 5 に示すように結果を表示し、認識結果の保存や候補となる5文字の中から適切な認識結果を選択することも可能である。その後、古典籍画像中のくずし字を再びタッチすることで、認識結果を何度でも参照することができる。また、認識結果および検出したくずし字の座標は、テキストデータとして翻刻された対象の画像を選択することで、画像と同時に読み込まれ、いつでも再参照することができる。

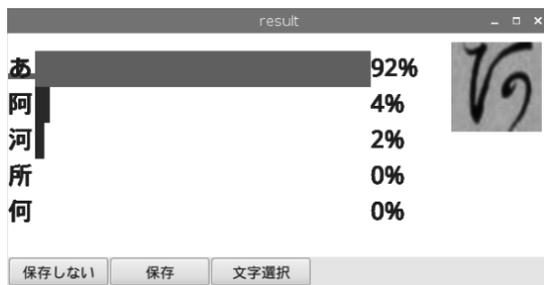


図 5 開発システムによるくずし字認識結果の表示例

認識を行った文字および検出を行った文字の確認をする場合は、ディスプレイを2本指でタッチすることで、図 6 に示すように、認識を行った文字は翻刻されて緑色枠で、文字検出を行った結果の領域は青色枠で囲まれて表示される。

検出した領域に対してユーザーが編集を行いたい場合は、図 6 に示す翻刻画像の左下にある領域編集ボタンにより可能である。図 7 に示すように変更したい領域を領域がすべて緑色枠に収まるように選択する。その後、図 8 に示すように領域の再選択を行い、確定ボタンを押すことで任意の領域に変更することができる。



図 6 開発システムによるくずし字の自動抽出および認識の実行例(河本家住宅保存会・島根大学附属図書館所蔵 河本家古典籍「山水下」[13]より)

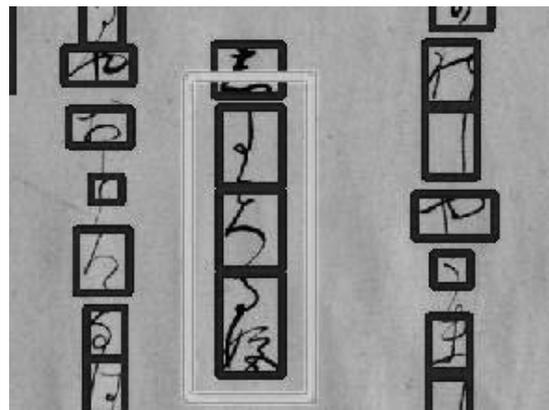


図 7 開発システムにおける編集領域の選択例

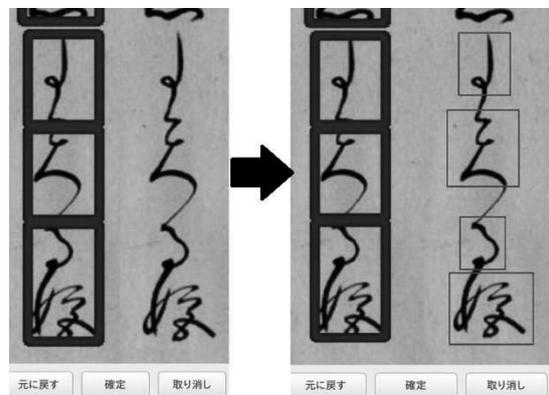


図 8 開発システムにおける検出領域の再選択の例

#### 4.5 テストデータに対する評価

ADEAC から提供のあった写本「山下水」[13](2,712文字)に対して本システムによるくずし字領域の自動抽出を行ったところ、62.2%の精度を得た。また、正しく抽出できた文字領域に対する認識率は、確信度最上位の文字に対して 57.3%、確信度が第5位までのすべての文字に対して 78.4%であった。

### 5. あとがき

本研究では日本語の歴史的典籍の自動翻刻を目的として、深層学習により畳み込みニューラルネットワークモデルにくずし字の認識を行わせ、それをくずし字 1 文字認識用の WWW アプリケーションとして実装した。結果として、高性能なハードウェアを利用しなくても十分な演算速度による翻刻が実現できることを明らかにした。デジタルアーカイブシステム ADEAC における歴史的典籍から時代や書き手、種類などのバリエーションを増やした字形データを学習データとして用いることによって、今後もさらなる認識率の向上に繋がることが期待される。

また本研究では、その深層学習モデルを利用して、カメラで撮影した歴史的典籍画像等からくずし字の自動検出および認識を可能とするシステムを開発した。Raspberry Pi のような小型で比較的安価なシングルボードコンピュータでも素早い認識を行わせることができた。小中学校などのネットワーク環境のない場所や普段モバイル機器を持ち歩かない高齢者の方々でも、くずし字に触れたい場面で支援ツールとして活躍することができるよう、操作性やくずし字領域の検出精度をより高めることが今後の課題である。

本研究の成果が発展することにより、近い将来、翻刻作業に人間を必要としなくなるという指摘もあるが、例えば機械翻訳技術が急速に発展している現在においても「翻訳」という職業はなくなるように、歴史的典籍を持つ古人の心を伝えるためには、例えば文学研究者の力が必要となると考えられる。研究者のみならず一般の人々が歴史的典籍を判読することを支援することで、海外における日本の歴史的典籍の利用価値を高め、それらに記された知識の遺産を有効活用することを促すことができると期待される。そのためにも本研究が果たす役割は少なくないと考えられる。

### 参考文献

[1] 国文学研究資料館. “歴史的典籍に関する大型プロジェクト”. <https://www.nijl.ac.jp/pages/cijproject/>, (参照 2015-10-14).

[2] T. Clanuwat, M. Bober-Irizar, A. Kitamoto, A. Lamb, K. Yamamoto, and D. Ha. “Deep learning for classical Japanese literature”. arXiv:1812.01718, (参照 2019-07-03).

[3] 早坂太一, 大野互, 加藤弓枝, 山本和明. “深層学習による変体仮名の翻刻および WWW アプリケーション開発の試み”. 人文科学とコンピュータシンポジウム論文集, 情報処理学会シンポジウムシリーズ, vol.2016, no.2, pp.7-12.

[4] 人文学オープンデータ共同利用センター. “日本古典籍字形データセット(国文研所蔵・CODH加工)”. <http://codh.rois.ac.jp/char-shape/>, (参照 2019-06-07).

[5] デジタルアーカイブシステム ADEAC. <https://trc-adeac.trc.co.jp/>, (参照 2019-07-12).

[6] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, “Caffe: Convolutional architecture for fast feature embedding.”. arXiv:1408.509, (参照 2015-11-12).

[7] 岡田一祐. “『和翰名苑』仮名字体データベース”. <https://kana.aa-ken.jp/wakan/>, (参照 2016-08-16)

[8] 人文学オープンデータ共同利用センター. “日本古典籍データセット(国文研所蔵)二十一代集”. <http://jcbv.nii.ac.jp/oa/NIJL0-1/items/NIJL0002.zip>, (参照 2016-07-25).

[9] 法書会編. “五體字類”. <http://www.let.osaka-u.ac.jp/~okajima/PDF/5tai/>, (参照 2015-11-12).

[10] 人文学オープンデータ共同利用センター. “源氏物語: 日本古典籍くずし字データセット”. <http://codh.rois.ac.jp/char-shape/book/200003803/>, (参照 2020-01-18).

[11] 京都女子大学・京都女子大学図書館. “淀藩土上月家文書 御着 城御当日御規式帳”, <https://trc-adeac.trc.co.jp/WJ11F0/WJJS07U/2672055100/2672055100200030/mp100030>, (参照 2020-01-18).

[12] TensorFlow Lite, <https://www.tensorflow.org/lite?hl=ja>, (参照 2020-01-18).

[13] 河本家住宅保存会・手銭記念館・島根大学附属図書館. “山陰地域史資料アーカイブ 河本家古典籍 山下水”, <https://trc-adeac.trc.co.jp/WJ11F0/WJJS07U/3290515100/3290515100200010/mp900000>, (参照 2020-01-18).

## **一般講演 4**

**座長:加藤常員(大阪電気通信大学)**



## 無形民俗文化財のデータベース化のための情報の 体系的記述に関する検討～浜松市における～

### A Study on Systematic Description of Information for Database of Intangible Folk Cultural Properties -In Hamamatsu City-

杉山岳弘

Takahiro Sugiyama

静岡大学情報学部, 静岡県浜松市中区城北 3-5-1

Shizuoka University, 3-5-1, Johoku, 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。今回、無形民俗文化財の祭りに関するデータベースに掲載する情報を拡張するために、民族学の分野において体系化されたコードとして、人間の行動・習慣・生産に関わる資料を分類するためのコードである文化項目分類(OCM:Outline of Cultural Materials)を参照し、どの程度適合するかを検証する。また、浜松市の無形民俗文化財「西浦の田楽」「川名のひよんどり」「横尾歌舞伎」「遠州大念仏」「つなん曳き」について情報の記述を試み、無形であるゆえの問題点を明らかにしていく。

**Abstract:** Hamamatsu City has more than 300 festivals, rituals, and events that have been handed down since ancient times, including intangible folk cultural properties designated as countries, prefectures, and cities. In collaboration with Hamamatsu City, our laboratory began collecting information about the festival from 2014, and as a result, built and released a festival database in 2017. This time, as a systematic code in the field of ethnology, a code for classifying materials related to human behavior, customs, and production was used to expand the information contained in the database on the festival of intangible folk cultural properties. Refer to OCM (Outline of Cultural Materials) and verify the degree of conformity. In addition, we tried to describe information about Hamamatsu City's intangible folk cultural properties, "Nishiure no Dengaku", "Kawana no Hiyondori", "Yokoo Kabuki", "Tsunan Biki", and "Enshu Dainen Butsu".

**キーワード:** 無形民俗文化財, 民俗芸能, HRAF, 文化項目分類

**Keywords:** intangible folk cultural properties, traditional folk entertainment, HRAF, outline of cultural materials

#### 1. はじめに

文化庁は、平成30年度の文化財担当者連絡会議にて文化財保護法改正について発表している[1]。この趣旨の中で、「過疎化・少子高齢化などを背景に、文化財の滅失や散逸等の防止が緊急の課題であり、未指定を含めた文化財をまちづくりに活かしつつ、地域社会総がかりで、その継承に取り組んでいくことが必要。」と述べている(なお、この改正は平成31年4月1日から施行されており、詳しくは[2]に公開されている)。この改正では、未指定を含めた文化財までを範囲として、活用と継承の必要性を述べている点で大きな拡張と言える。本研究では、民俗文化財のうち風俗慣習(儀礼、年中行事、祭礼など)と民俗芸能(念仏踊、歌舞伎、神楽、

田楽など)、一般に「祭り」と称される中で行われる行事を対象としていく。

無形民俗文化財における未指定まで民俗芸能「祭り」の継承・活用を考えると、まずは地域における民俗芸能の全体の把握とデータベース化が重要となる[3, 4]。データベースとして、構造化の難しさや、どういった情報が必要であるかの議論はいくつかなされている[5, 9, 13]。本研究の立場としては、祭りに関する普遍的な情報や、目的に応じた必要な情報を保存するという立場ではなく、人の日々の生活の中で行われていく、無形であるがゆえに、時代の移り変わりに応じて、年々変化していく「祭り」を捉え、記録していきたいと考えている[5]。

本研究では、平成 28 年度の調査で、浜松地域の民俗芸能、主に「祭り」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[6]。

祭りに関するすべての情報を捉えることは難しいが、研究者が調べまとめた資料や関係者のインタビュー、様々な人が撮影した祭りの写真・映像といった断片的に得られる情報を、体系的にまとめあげ、参照できるようにすることはアーカイブとしても大きな価値がある。

本研究では、西尾らによる文献[5]の検証のため、浜松市の無形民俗文化財「西浦の田楽」「川名のひよんどり」「横尾歌舞伎」「遠州大念仏」「つなん曳き」について情報の記述を試み、無形であるゆえの問題点を明らかにし、体系化を行う上で必要となる課題を検討する。さらに、民族学の分野において体系化されたコードとして、人間の行動・習慣・生産に関わる資料を分類するためのコードである文化項目分類(OCM:Outline of Cultural Materials)を参照し、民俗芸能についての適合性について検証する。

## 2. 体系的記述に関する検討の概要

無形民俗文化財「祭り」に関わる情報は、無形ゆえに形がなく、さらに入手できる情報は断片的であることが多く、全体像を捉えることは難しい。また、口承により継承されてきた祭りも多く、情報の確実性に関わる問題も潜在している。そのため、一度、途絶えてしまうと、断片的な資料と経験者の記憶に頼らざるを得ず、その復活は困難を極める。そのため断片的で不確実な情報を、体系的に捉えることで、全体像として俯瞰して理解しやすくなる仕組みが重要となってくる。

### 2.1 神澤のおくないの事例

近年、多くの伝統芸能が消滅している。浜松市においては、浜松市天竜区佐久間町浦川に伝わる「浦川歌舞伎」が 2019 年 9 月を最後に終了している。1962 年に一度途絶え、1989 年に有志により復活させ 30 年ほど継続していたが、地元の人口減少と高齢化により継続を断念した。最近だと、2020 年 1 月において宮城県石巻市雄勝町名振で、約 240 年にわたり行われてきた県指定無形民俗文化財「おめつき」も中止となっている。

静岡県浜松市天竜区神沢では「神沢のおくない」が復興を果たして現在も継続してきている。神沢のおくないは、昭和 30 年代後半に完全に途絶えた

が、約 50 年を経て、平成 21 年に復興した。現在も、毎年 1 月 5 日に、中学生による舞がおくない堂で披露されている。神澤おくない同好会事務局の石野重利氏は、この祭りの復興にあたり、舞の所作や笛の演奏方法を記載した「神澤おくないの手引き」を作成した。

この手引の作成のために石野氏は、神沢のおくないに関する研究資料[7]、かつての舞手の意見、自身の記憶を元にしたという。ただし、資料は非常に少なく、ほとんどは舞手の意見を参考にしたという。また、衣装の記録がなかったため、類似した祭りを参考にしたという。以上のことから、この復興が可能であった理由は、舞手から当時の情報を得られたこと、周辺の類似した祭りの情報があつたことの 2 つであることが言える。

この事例が示すのは、例え断片的であっても、祭りの情報を分かるうちに保存しておくことの重要性である。さらに、類似した祭りの情報を一体として体系的に俯瞰していくための仕組みが重要となってくる。

### 2.2 断片的な情報の記述に関する検討

本研究の先行研究である文献[5]をもとに、祭りのデータ構造を用いる。基本となるデータ構造を、本稿末尾の表 1 にまとめておく。祭りの変遷を記述する仕組みは、多くの情報を集めることができるある年(とし)の情報をマスターデータとし、これを基準として、その年の前や後の年毎に祭りの変化した部分をデータとして記述する[5]。

データを記述する年について、手に入れられる資料においては、資料の発行年や祭りの実施年といった複数の年があるが、祭りの実施年を基本として、対応する年のデータとして記述していく。多くの資料は、情報が断片的であるので、埋められる項目は限られており、まばらとなる。

この枠組みを使って、浜松市の 5 つの祭り：西浦の田楽、川名のひよんどり、横尾歌舞伎、遠州大念仏、つなん曳きに対して、断片的な情報を組み合わせ記述を試み、課題を明らかにしていく。実際に祭りに関して参照した資料(一部抜粋)は次の通りである。

#### 西浦の田楽(国指定重要無形民俗文化財)

- 鑑賞の手引(冊子/1995 年/浜松市発行)
- 西浦田楽紹介サイト(Web/2018 年/杉山研究室)
- 西浦のまつり(写真集/1970 年/須藤功)
- お祭りデータベース(Web データベース/2017 年/杉山研究室)
- 遠江・山と里の民俗(DVD/2015 年/浜松市中山間地域の文化遺産活用実行委員会)

- 水窪の民俗(書籍/2012年/遠州常民文化談話会)
- 水窪町史 下(書籍/1982年/水窪町史編纂委員)
- 撮影(映像・写真/2017年/杉山研究室)

#### 川名のひよんどり(国指定重要無形民俗文化財※ 遠江のひよんどりとおくない)

- 静岡県の祭ごよみ(書籍/1990年/静岡県民俗学会)
- ガイドブック「遠江のひよんどりとおくない」  
(冊子/2011年/遠江のひよんどりとおくない連絡協議会)
- 静岡県引佐町川名 福満寺薬師堂祭礼「川名のひよんどり」(冊子/作成年不明/川名ひよんどり保存会・引佐町教育委員会)
- インタビュー調査資料(テキスト・音声/2016年/杉山研究室)
- 遠江・山と里の民俗(DVD/2015年/松市中山間地域の文化遺産活用実行委員会)
- 撮影(映像・写真/2016~2019年/杉山研究室)

#### 横尾歌舞伎(静岡県指定無形民俗文化財)

- 静岡県指定無形民俗文化財 横尾歌舞伎(冊子/2015年/横尾歌舞伎保存会)
- インタビュー調査資料(テキスト・音声/2016年/杉山研究室)
- 遠江・山と里の民俗(DVD/2015年/浜松市等)
- 地芝居のある村(DVD/2015年/浜松市中山間地域の文化遺産活用実行委員会)
- 撮影(映像・写真/2016・2017年/杉山研究室)

#### 遠州大念仏(浜松市指定無形民俗文化財)

- 遠江・山と里の民俗(DVD/2015年/松市中山間地域の文化遺産活用実行委員会)
- 調査資料(テキスト・写真/2015年/杉山研究室)
- 静岡県の祭ごよみ(書籍/1990年/静岡県民俗学会)
- 広報はままつ特集号(冊子/1999年/浜松市)
- 浜松市ホームページ内「遠州大念仏」  
(Web/2013年/浜松市)
- 遠州大念仏保存会ホームページ「所属団体」  
(Web/2018年/遠州大念仏保存会)
- 撮影(映像・写真/2017年/杉山研究室)

#### つなん曳き(浜松市指定無形民俗文化財)

- 遠江・山と里の民俗(DVD/2015年/松市中山間地域の文化遺産活用実行委員会)
- インタビュー調査資料(テキスト・音声/2016年/杉山研究室)
- お祭りデータベース(Webデータベース/2017年/杉山研究室)
- 浜松鈴鈴2016年第1号(冊子/2016年/浜松鈴鈴)
- 撮影(映像・写真/2016・2017年/杉山研究室)

## 2.3 体系的に俯瞰する仕組みの検討

神沢のおくないの復興のときに、石野氏は類似した祭りを参考にしている。このような類似した祭り同士を結びつけて関連する情報を俯瞰するためには、単なるキーワード検索では不十分で、メタデータの設計が不可欠である。無形民俗文化財においていくつか検討がされている[8,9]。今回は、直接的には踏み込まず、体系的な分類のための枠組みについて検討する。

文化人類学・民族学の分野では、HRAF(Human Relations Area Files:フラーフ)[10]という「世界中の様々な民族の社会や文化について書かれた文献(単行本、論文等)を、地域・民族別に集め、それらすべてのページの内容を専門家が独自の分類方法を使って分析したファイル資料」(Webページ[11]より引用)がある。これを利用して、民俗芸能の様々な資料に対して分類することができれば、体系的に利用できるようになると考えられる。今回は、このHRAFの分類で用いられているOCM(Outline of Cultural Materials:文化項目分類[12])について、民俗芸能に関わる項目がどの程度適合できそうかを調査する。

## 3. 情報の記述に関する検討

先行研究[5]で提案している、今回入力に用いるデータ構造を本稿末尾の表1に示す。データ構造に対する情報の記述に関する検討では、「西浦の田楽」「川名のひよんどり」「横尾歌舞伎」「遠州大念仏」「つなん曳き」の5つの祭りを対象に、実データを入力する。また、入力にあたっては、このデータ構造[5]の検証と同時に、祭りの経年変化を継続的に記録していくことを考慮して、祭りに関する調査補助を1年間実施した一般の人に、これまで研究室で収集した資料をもとに行ってもらった。

本データ構造[5]では、祭りの変遷についても記述をするために、充実した情報を持つ資料がある祭りを実施した年をマスターデータとし、そのマスターデータの年を基準として時を経て、もしくは遡り「変化した部分」について、その年を追加してデータとして入力する。また、必要に応じてデータ項目を追加、整理を行った。

以下については、入力者の報告をもとに作成してある。

### 3.1 西浦の田楽

西浦の田楽は、静岡県浜松市天竜区水窪町で行われている祭りで、観音堂境内にて旧暦1月18日の月の出から翌日の日の出まで、夜を徹して舞を主とした神事が行われる。その歴史は約1300年と

長く、国指定重要無形民俗文化財となっている。

1989年に水窪教育委員会（現在は浜松市発行）が発行した鑑賞の手引をマスターデータとした。西浦の田楽は旧暦の決まった日に行われるため、毎年開催日が変わる。その変遷を記述するため、直近4年間の開催日の情報を入力した。

#### 入力において問題になった点

資料中に図や画像があった場合に、どうデータの入力を行うかが問題になった。今回は備考欄に図等がある旨と該当ページを入力することで対応した。

今回マスターデータとした鑑賞の手引には、開催場所の住所や地図座標、文化財区分や伝承状況等についての記載がなかった。こういった、マスターデータに必要と思われるが不足している情報を、一部別の資料から抜き出した。

行事の区切り方について、日別、舞や演目別などいくつかの視点がある。また西浦の田楽については、舞の演目が47演目と多いため、今回のデータベースとは別に、舞を一体とした別のデータベースが必要と感じた。

西浦の田楽では演目ごとに使われる楽器や道具が異なる。楽器は楽器の項目に、道具は衣装の項目に入力したが、リストでは関連性が失われてしまい情報が減る。舞ごとにまとめた方が、情報を維持できると考えられる。

#### 課題

マスターデータを複数の資料から作成することの検討（祭りによっては断片的な情報を持った複数の資料が存在する場合もあると考えられる）、どこまで詳細な情報を入れるか、また他のデータベースとの対応付けの方法も検討が必要である。

### 3.3 川名のひよんどり

川名のひよんどりは、同市内北区引佐町で行われている祭りで、火踊りがなまり伝えられたのが、ひよんどりの名称の起こりといわれ、400年以上前から続く祭りである。厳冬の中、裸の若者が川で身を清め、堂の入り口で火に当たりながら揉み合いをする姿は圧巻であり、静岡県は無形文化財に指定されている。

マスターデータには、静岡県民俗学会が発行している書籍「静岡県の祭ごよみ」に収録されているデータを入力した。

#### 入力において問題になった点

マスターデータを入力後、より詳細で網羅されている資料が分かった場合に、今回はマスターデータを基準に他の資料については差分を入力した。もしマスターデータとなる資料を変更した場合は、新

しいマスターデータを基準に差分を入れるとなると、今まで入力したデータを全部書き直す必要がある。より詳しい資料が出てきても、あくまで最初に定めたマスターデータを基準にすべきと考えられる。

西浦の田楽同様、説明画像の扱いが現状の枠組みでは困難であった。各舞や面、行事、エピソードなど各項目に画像の項目の追加が必要と考えられる。

作成年不明の資料がいくつかあった。一次資料としては扱えなくなるが、推定で書くことになる。

そして、行事と舞の対応付けをはじめ、資料の内容をどの項目に記述するか判断が難しかった。

#### 課題

西浦の田楽同様、マスターデータの基準の再検討が必要と考えられる。行事と舞といった複数項目の対応付けや、資料中の情報がどの項目に入るか判断するための一定の基準の検討が必要と思われる。特に祭りに詳しくない人がデータ入力を行うことも想定した基準の検討が必要である。

### 3.4 横尾歌舞伎

横尾歌舞伎は静岡県指定無形民俗文化財の祭りであり、同市内北区引佐町で行われている。発祥は定かではないが、保存する資料などから200年前より盛んに上演されていたものと考えられる。この祭りの特徴は、役者・太夫・三味線弾きから、振付・着付・床山・大小道具・勘亭流の書体によるポスター作りに至るまで全て地域の人達の手で賄われていることである。

マスターデータには、平成27年度の公演資料と、同年に作られたと思われる祭りの説明資料を使用した。

#### 入力において問題になった点

ときに内容が一緒くたになっている資料があり、どの項目を入力するかは適宜主観で判断しているのが現状である。

また、組織に関する情報の記述が少なかった。継承・活用という点では、組織に関する情報は重要であるが、公式な資料はあくまで祭りの紹介に重きが置かれており、内部のことまで触れられていることは少ない。内部のことはあまり外に出したくないという印象がある。

#### 課題

川名ひよんどり同様、祭りに詳しくない人でもある程度判断できるための指針が必要と思われる。

### 3.5 遠州大念仏

遠州大念仏は市指定無形民俗文化財の祭りである。おこない自体は遠州地方の郷土芸能のひとつ

で、初盆を迎えた家からの依頼でその家の庭先で大念仏を演じる。市内全域で約70の組が保存会に所属しており、各地で祭りを行っている。一般向けには例年7月15日に中区の犀ヶ崖にて大念仏が行われている。

1999年に市の広報課が公開した「はままつ歳時記 広報はままつ特集号」をマスターデータとした。

#### 入力において問題になった点

遠州大念仏自体は、市内全域複数箇所で行われているが、公式な資料としては、一般向けに犀ヶ崖で行われているものしか見当たらなかった。

#### 課題

複数の場所で行われている祭りの記述の枠組みの検討が必要である。

### 3.6 つなん曳き

つなん曳きは、市指定無形民俗文化財の祭りである。天竜区春野町犬居地区で行われており、その起源は、過去にこの地域で大洪水が起きた際に、2匹の龍が村を守ったという伝説が基となっている。毎年祭り当日に、竹などで大きな龍を作り、それを地元の人達が担いで町内を練り歩くというものである。

この祭りでは公式と言える資料が見当たらず、過去に筆者らが祭りの主催者にインタビューを行い作製した地域情報誌「浜松鈴鈴 2016年第1号」をマスターデータとした。

#### 入力において問題になった点

遠州大念仏等にも言えるが、指定文化財の規模が小さくなるにつれ、一次資料に適した資料がない場合が多い。

#### 課題

指定文化財等何らかの基準ごとに、入力の方針を変えることも検討したい。

### 3.7 情報の状態についての考察

今回以下の3つが挙げられた。

・必ず変化する

→ある年の情報を追加すると必ず入力するもの参照年度や参考資料、開催日等がそれにあたる。

・変化することがある

→状況が変化した時に入力するもの伝承状況や行事、アクセス情報等。

・ある年だけ項目に値が入る

→祭りの停止や中止、備考等。

#### (1) 空欄の意味付けの必要性

必ず変化するものは空欄になることはないが、問題となるのは空欄の項目である。年によって断片的な情報しか入手できないため、空欄には空欄

となった意味づけが必要となる。空欄が空白であるという情報は、変化がなかった場合と、情報そのものがなかった場合の両方の可能性があり、その情報の状態を表す情報を記録しておく必要がある。

#### (2) 入手状況による情報の状態

また、入手できた情報の状態は、研究者により調べられた資料から、教育委員会や有志によって作られて資料、さらに、祭り当事者が経験した情報、当事者が先代から聞いた情報、その地に残る言い伝え、といった真偽性のような情報を記録しておく必要がある。

#### (3) 複数の情報源の整合性の取り方

複数の情報源があるとき、異なる記述や、場合によってはどちらかが間違っている可能性のある記述があり、その整合性が問題となる。これについては西浦の田楽について当事者からインタビューを行い、検証を行っている研究がある[13]。

#### (4) 正確な場所の特定

祭りはある地域で行われるが、一箇所でおこなわれるわけではなく、複数の場所でそれぞれの行事が行われる場合もあるので、その記述を検討する必要がある。また、「祭り」の括り方によっては、複数の地域で、地域によっては別の日に、同じ祭りが行われる場合もある。これを地域に分けることで曖昧無く記述できるが、祭りの同一性を記述する手段が必要である。

#### (5) 日にちや期間に関する情報について

日にちについての問題は、すでに議論がされている[4,9]。また、資料についても作成された日付が曖昧なものもあり、また、ある人がある人から聞いたとか、伝承で江戸時代に起こった事件など、期間が曖昧なものもある。これでもできるだけ曖昧無く記述すべきである。これについても仕組みを提案している論文がある[14]。

#### (6) 情報の入力について

祭りの記述の対応付け、情報の関連性の記述、同じ項目に複数の資料からの情報があつたときの判断基準など、検討すべき課題が多い。

#### (7) 情報の入手について

今回データ入力を行った5件の祭りについては、市レベルから国レベルまで、何らかの文化指定を受けているものに絞って入力を行った。しかし、文化庁の考え方では、未指定を含めた文化財までを範囲としている。過去に筆者らが行った、浜松市内全域の祭り全点調査では、未指定を含め300件以上の祭りが確認された。今回の5件の中だけでも、規模の違いによって、残っている資料の量や内容の粒度に大きくばらつきがあつたので、幅広く記録に残すためには、祭りの規模ごともしくは何らかの視点でデータ項目を複数パターン検討する必

要があると考えられる。また、継承・活用という点では、現存する資料から抜き出していくだけではなく、定義したデータ項目に沿って、祭りの主催者に積極的に聞き出していくことも必要と考えられる。

#### 4. 文化項目分類の民俗芸能への適合性

項目の語彙として文化項目分類を検討した。文化項目分類の民俗芸能への適合性の検証については、表1のデータ構造の各項目について、文化項目分類におけるコードに対して、直接的(語彙の一致)・間接的(概念の一致)に関係するものを対応付けていき、どの程度対応付けられるかを試みる。実際のところ、世界の民族文化を対象としていることに対して、本研究の日本の民俗芸能に関しては対応付けられる記述は少なく困難であった。以下は具体例である。

- ・日付に対応するコード/名称が見当たらなかった(上述が理由と思われる)
- ・神輿や屋台といった地域独自のものは見当たらなかった(道具や建築の中に合いそうなものがあるかと探したが、やはり見当たらなかった)
- ・項目名としては合いそうだが、意味的に違う場合がある(例:206.04/組織と人員→「組織」と置き換えられるかと思っただが、大項目である206は「電話と電信」。例3:41/道具と機器が使えそうと感じたが、工業活動に用いられる器物の種類であった)

しかし、いくつかは、本来の使い方として合いそうな名称はいくつかあった(例:別火に入ることの説明文に対して、264.07/食事行動に関する信仰や慣習など)。

文献・資料の体系的分類という観点から、HRAFについて調べていたが、祭りの体系的な扱いにはやはり民俗関係の辞典や資料から体系化を行い、枠組みとしてHRAFを参考すべきである。

#### 5. まとめ

本研究では、無形民俗文化財における民俗芸能の持つ情報の体系的な記述のため、過去に提案したデータ構造に、浜松市の5つの祭りについて実データの入力を行い、空欄の意味づけの必要性、複数の情報源の整合性の取り方、入手状況による情報の状態など課題を明らかにした。さらに、祭りを体系的に分類して俯瞰するため、文化人類学の文献・資料の分類で用いられる文化項目分類を、民俗芸能に当てはめ適合性を検証した。結果、現段階ではあまり有効でなく、無形民俗文化財の分野における項目分類をまとめる必要があることが分かつ

た。今後は、さらに祭りに関する実データの入力を増やし、情報の状態を記述することができるデータベース・システムを設計・実装・検証を進めることと、民俗芸能の項目分類の作成を検討していく。

#### 謝辞

本研究の一部は科研費基盤研究(C)19K12712の助成を受けたものである。

#### 参考文献

- [1] 文化庁, 「文化保護制度の見直しについて」, [https://www.bunka.go.jp/seisaku/bunkazai/pdf/r1414902\\_01.pdf](https://www.bunka.go.jp/seisaku/bunkazai/pdf/r1414902_01.pdf) 2019年1月, (参照2020-01-22)
- [2] 文化庁, 「文化財保護法及び地方教育行政の組織及び運営に関する法律の一部を改正する法律等について」, <https://www.bunka.go.jp/seisaku/bunkazai/1402097.html>, 2019年1月, (参照2020-01-22)
- [3] 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2020-01-22)
- [4] 杉山岳弘, 戸田剛, 太田好治, 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, 第23回公開シンポジウム「人文科学とデータベース」, pp. 9-16, 2018.3
- [5] 西尾美沙季, 杉山岳弘, 「無形民俗文化財「祭り」の保存と継承に必要な情報とその変遷をアーカイブ化するための記述方法の検討」, 情報処理学会第78回全国大会, 4ZB-05, 2016.3
- [6] 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2020-01-22)
- [7] 新井恒易: 中世芸能の研究, 新読書社, pp. 526-560 (1970).
- [8] 彦坂和里, 杉山岳弘, 「西浦田楽」の演目映像デジタルアーカイブ化に向けたメタデータの検討, 人文科学とコンピュータシンポジウム論文集(じんもんこん2019), pp. 71-76, 2019.12
- [9] 佐藤いつみ, 高久雅生, Linked Dataによる無形民俗文化財情報の構造化の試み, 人文科学とコンピュータシンポジウム論文集(じんもんこん2019), pp. 245-252, 2019.12
- [10] 稲葉洋子, HRAFと国立民族学博物館所蔵図書の種類, 情報の科学と技術, 第58巻, 第2号, pp. 64-70 (2008)
- [11] 国立民族学博物館, Human Relations Area Files (略称 HRAF フラーフ), <https://www.minpaku.ac.jp/research/sharing/library/search/hraf> (2020-01-22 参照)
- [12] G. P. マードック(他編), 国立民族学博物館(訳), 文化項目分類, 国立民族学博物館, (1989)

- [13] 彦坂和里, 杉山岳弘, 「西浦の田楽」の演目映像デジタルアーカイブ化に向けた検討 ～演目に関する複数年代の文献調査から見える課題～, 第24回公開シンポジウム「人文科学とデータベース」, pp. 9-15, 2019.3
- [14] 関野樹, 時間名による時間参照基盤の構築-Linked Dataを用いた期間の記述とリソース化-, 人文科学とコンピュータシンポジウム論文集 (じんもんこん 2019) , pp. 267-280, 2019.12



古天気・作況・死亡データベース構築の構想  
-18・19世紀の東北地方における死亡危機の要因解明に向けて-  
**Necessity for the database system on historical weather, rice harvest,  
and mortality to investigate the cause of the mortality crisis in  
northeastern Japan, the 18th and the 19th century**

川口 洋

Hiroshi Kawaguchi

帝塚山大学 文学部, 奈良県奈良市帝塚山 7-1-1

Tezukayama University, 7-1-1, Tezukayama, Nara, Nara

**概要:** 北関東以北では18世紀初頭から100年以上にわたって人口が減少した。黎明期の人口史家は、人口減少の主要因を宝暦、天明、天保期に発生した冷害にともなう凶作と考えた。1960年代末に歴史人口学が導入されると、死亡危機が頻発しており、既知の凶作年は死亡危機の一部に過ぎないことが確認された。しかし、人口分析の対象となる史料が保存されている集落の近隣で天気や作況などを記録した史料が発見されていないため、死亡危機の要因は未だ解明されていない。本稿では、筆者が陸奥国会津郡金井沢村で再発見した古文書史料をもとに、既開発の寺院「過去帳」分析システムを用いて、死亡危機と作況との関係を展望するとともに、古天気、作況、寺院「過去帳」データベースを基幹とする死亡危機の要因分析システム構築の必要性について述べる。

**Abstract:** The population in north-eastern Japan had dropped for over hundred years as of the beginning of the 18th century. Some historians pointed out that the main cause of population decrease must be the poor harvests resulted from cool summer in 1755, 1783, and the 1830s. Historical demographers have confirmed that there were many mortality crisis years in addition to the poor harvest years. As we cannot find the historical documents recording historical weather, rice harvest, and mortality in nearby village, we have not clarified the causes of the mortality crisis yet. In this paper, I would discuss the relationship between mortality crisis and rice harvest in the Village of Kanaizawa, Aizu County in the Province of Mutsu. I would also propose to construct a database system for investigating the causes of mortality crisis. The system will be composed of the database for historical weather, the database for rice harvest, and the database for the Buddhist temple death registers.

**キーワード:** 死亡危機, 古天気, 作況, 寺院「過去帳」, 飢饉

**Keywords:** mortality crisis, historical weather, rice harvest, Buddhist temple death register, famine

## 1. はじめに

徳川幕府による人口調査によれば、北関東以北の人口は1721年から約100年にわたって減少を続けたのち、ゆるやかに回復を始めた。1840年代以降の持続的人口増加は、多産多死の伝統社会から少産少死の産業化社会への助走を示す指標の一つと理解されている。

黎明期の人口史家は、農業生産が不安定で商品作物の導入が遅れた東北地方では、宝暦・天明・天保の冷害・凶作・飢饉で多くの餓死者が犠牲となり、領主の苛斂誅求に喘いでいた民衆が随胎や間引きといった出生制限を行った結果、江戸時代後半の人口が減少したと考えた[1][2]。このような認識は、史料を離れて独り歩きを始め、東北地方だけではなく、江戸時代

後半の窮乏した民衆像・歴史像を形成するうえで大きな影響を及ぼした。

「宗門改帳」や寺院「過去帳」に歴史人口学の研究方法が適用されると、18・19 世紀の東北地方に生きた人々の生涯が、次第に浮き彫りにされていった[3][4][5][6][7][8][9]。黎明期の人口史家が描いた民衆像、歴史像に再検討を迫った研究成果のなかで、粗死亡率が極めて高くなる死亡危機は頻繁に発生しており、既知の宝暦・天明・天保期の凶作・飢饉年は死亡危機の一つに過ぎないことが確認された[10]。

歴史人口学の研究成果が蓄積されてきたにもかかわらず、死亡危機の実態が不鮮明であるのは、次の理由による。①天気と作況の長期変動を記録した史料が発見されていないため、冷害・凶作にともなう死亡危機と冷害・凶作と連動しない死亡危機を区別することが困難であり、凶作の程度と死亡指標との関係についても検討できない。②史料的制約に阻まれて、作物の生産・流通・消費と死亡指標との関係が十分解明されていない。③人口規模が小さな村を対象とした人口分析の結果が、どの地域的範囲を代表する事例か明示しにくいいため、②に関わる隣接分野の研究成果と接続することが容易ではない。

①については、1 カ村の天気、作況、死亡者を長期間記録した史料の発見が喫緊の課題であった。筆者は、陸奥国会津郡金井沢村の名主家が、100 年以上にわたって天候や農作業を記録した「農業日記」、1 坪から収穫された粃の収量を記録した「作毛位付帳」、人口動態を記録した「人数増減差引之覚」、および死者を供養した寺院「過去帳」を再発見することができた。

本稿では、金井沢村の史料と既開発の寺院「過去帳」分析システムを用いて、天明 3(1783)年の冷害にともなう大凶作を挟む 18 世紀後期の凶作と死亡危機の関係について展望する。さらに、古天気データベース、作況データベース、および寺院「過去帳」データベースを基幹とする 18・19 世紀における死亡危機の要因分析システムの必要性について述べる。

## 2. 史料

18・19 世紀の東北地方における死亡危機の要因分析システムの基礎史料となるのは、筆者が再発見した陸奥国会津郡金井沢村(現在の福島県南会津郡南会津町)の名主家で作成された毎日の天気が記録されている「農業日記」、1 坪から収穫した粃の収量を記録した「作毛位付帳」、および周辺の寺院で死亡者を供養した寺院「過去帳」である。

現在のところ、金井沢村は、1 カ村で 100 年以上にわたって天気や稲の作況と死亡指標との関係を検討することのできる日本唯一の村とみられる。

### (1) 「農業日記」

「農業日記」(奥会津博物館架蔵、室井家文書)には、明和 9(1772)年から大正 9(1920)年に至る 149 年間のうち、散逸した明治 2(1869)年から明治 39(1906)年などを除く 93 年分の毎日の天気、農作業、農業経営の詳細が記録されている。

### (2) 「作毛位付帳」

「作毛位付帳」(奥会津博物館架蔵、室井家文書)には、宝暦 9(1759)年から大正 5(1916)年に至る 158 年間のうち、散逸した天保 12(1841)年から明治 5(1872)年、明治 11(1878)年から明治 15(1882)年を除く 118 年分の 1 歩(1 坪)の水田 2~5 カ所から収穫された粃の収量が掲載されている。本史料には、歩刈を行った年月日、小地名、田(地)主、品種、株数、粃の重量、および粃の容積が記録されている。煙草、麻、大豆、小豆、蕎麦、粟、稗、芋、大根といった作物の作況が、上中下で判定されている年もある。

歩刈を行った水田の小地名や田主が多様であるため、歩刈の実施主体は金井沢村とみられる。歩刈の対象となる水田の選定に関する史料は未見であるが、中部地方の膨大な坪刈帳の分析にもとづいた、「坪刈りは一筆の水田でも中分の作柄部分を試刈し、しかも村内水田の地位別に数筆以上にわたって実施しており、一坪収量はその年におけるムラ(区)の標準稲作生産量に近似する数値となるよう配慮されている。...(中略)...ムラ(区)を構成する農民層の間で稲作技術の

階層間格差が著しいような生産力段階では、少なくとも一坪収量の計測といった方法が登場することはないのであり、村内において稲作技術の平準化が進行していることの証左である[11]との指摘は卓見である。

本史料は、100年以上にわたる歩刈の結果が記録されている東北地方唯一の貴重な史料である。

### (3) 寺院「過去帳」

高野村、上塩沢村、下塩沢村、大豆渡村、金井沢村、福米沢村、黒沢新田村、針生村、および木地小屋から構成される高野組に位置する寺院は A 寺と B 寺である。両寺院には、17 世紀以降の寺院「過去帳」が保存されており、戒名、死亡年月日、俗名または施主との続き柄、居村などが記録されている。なかには、被葬者の死亡年齢、生年月日、死亡地、出身地などが記されている場合もみられる。

両寺院の檀家は、高野組の全村と高野組に隣接する川嶋組中荒井村に分布する。金井沢村、福米沢村、大豆渡村、針生村、黒沢新田村では、神職や修験を除くほぼすべての家が A 寺または B 寺の檀家であるため、5 カ村におけるほぼすべての死亡者を両寺院の寺院「過去帳」で確認することができる。

## 3. 天明 3 (1783) 年の異常気象

「天明三年 農業萬日記」(室井家文書、4250)によれば、1783 年の金井沢村で降雨が記録されている日数は、6 月に 10 日(不記載 6 月 20 日、27-30)、7 月に 4 日(不記載 7 月 1-7 日、17 日)、8 月に 18 日、9 月に 10 日である。7 月 16 日・28 日・30 日、8 月 8 日・30 日・31 日、9 月 1 日・22 日、および 10 月 4 日に「サムシ」と記されている。

会津藩士・田村三省が書き残した「孫謀録」<sup>そんぼうろく</sup>には、「六月十五日、小雨降東風日々吹て冷気を催す。…(略)…十六日、十七日、東風吹[12]」と、7 月 14 日から 16 日に会津盆地で寒冷な東風が吹いたことが記録されている。東風は、オホーツク海高気圧から本州北部太平洋側に吹き出す寒冷な北東気流(ヤマセ)とみられる。「天明三年凶作」(室井家文書、3248)には、

「六月十日頃、別而不気候。田嶋御祭り之節杯、拾或者わた入の上へかたひら杯着候テ参詣いたし候躰<sup>ニ</sup>而、以之外さむし」とある。7 月 14 日に祇園祭で賑わう田嶋村周辺の百姓は、冬着である 袷<sup>あわせ</sup> や綿入れの上にかたひら<sup>かたひら</sup>帷子を重ね着して田出宇賀神社に参詣した。1784 年 7 月 14 日から 16 日まで、会津盆地と奥会津では、寒冷湿潤な東風の吹走にともなう異常低温と降雨が観察されていた。

奥会津では、1783 年 6 月中旬の梅雨入りから 9 月中旬の秋霖明けまで断続的に続いた長雨と浅間山の噴火にともなう 7 月 27 日から 8 月 4 日までの降灰が重複して、異常低温と日照不足に見舞われた。6 月中旬の田植えからおよそ 1 カ月を経た 7 月中旬から異常低温が記録されているため、1783 年は、稲の分けつ期に生育が遅れ減収となる遅延型冷害と生殖成長期に低温のため不稔となる障害型冷害が重複した複合型冷害に分類できる。

## 4. 18 世紀後半の作況

「作毛位付帳」には、118 年中 90 年分は 3 カ所、17 年分は 2 カ所、5 年分は 5 カ所、4 年分は 4 カ所、1 年分は 1 カ所の歩刈の結果が記録されている。明治期以降は上田、中田、下田の 3 カ所で歩刈が行われたが、江戸時代には上中下田の別が書かれていない年も多い。複数の水田で歩刈が行われて、上中下田の別が書かれている場合、例外なく最初に上田における歩刈の結果が書かれている。そのため、上中下田の別が書かれていない場合も、最初の記録は上田における歩刈の結果と推定される。

1759 年から 1799 年までの間、稲の品種は細葉(1759、1760、1795 年を除く)、歩刈した水田の小地名は柿下(1759、1760、1765、1766、1768、1769、1770、1878-1782 年を除く)、歩刈の実施日は、彼岸を挟む 9 月 19 日から 10 月 9 日である(1783 年を除く)。歩刈の対象となった水田の田(地)主は、加右衛門(1776、1777、1784-1799 年)である年が多い。彼岸前後の柿下周辺における上田で栽培された同品種の収量が記

録されているため、1759年から1799年の1坪から収穫された粃の総重量と容積を図1に示した。

「作毛位付帳」には、「天明三卯年 田方青作ニ而、當時歩刈可仕躰ニ無御座候」という記述とともに、粟・稗・蕎麦・菜大根・麻が下、小豆・芋・煙草が中、大豆が上と判定されている。「忠春日記」も、「粟、稗、そば等も実入りなくして、所により九月二日朝には、霜ふり申候[13]」と証言している。1783年は、例年歩刈を行う彼岸過ぎとなっても、歩刈ができる状態ではなかった。稲ばかりか、粟、稗、蕎麦も登熟せず、1783年9月27日に初霜が降りた標高700mを超える高所では、中下層の常食となる収穫直前の粟、稗、蕎麦まで壊滅的な不作となった。

1783年の歩刈は、彼岸前後に行われていた例年より1カ月以上遅れて、若松から派遣された検見役人かまどめの下で実施された。鎌止が厳命されたために雪のなかで11月6日に行われた検見歩刈の結果は、上田、稲草：ほそば、稲株：127株、粃：3合5夕(下見：2合5夕)、重量：32匁、下田、稲草：ほそば、稲株：118株、粃：2合5夕(下見：1合5夕)、重量：20匁と記録されている(「天明三年 御検見御歩刈書上帳 卯十月十二日 会津郡高野組金井沢村」(室井家文書、3251))。歩刈の結果、未曾有の不作が確定した。

1783年11月に上田1坪から収穫された粃の重量と容積は、「作毛位付帳」に記録されている118年間で最低となった(図1)。粃の重量は、天保4(1833)年の150匁、天保9(1839)年の260匁、明治35(1902)年・38(1905)年の270匁に遠く及ばず、1764年から1799年に至る36年間の平均重量482匁の1割にも満たない。粃の容積も、天保4年の6合5夕、天保9年・明治38年の9合5夕、天保6年・明治35年の1升到及ばず、1759年から1799年に至る41年間の平均収量1升7合7夕の2割程度である。

## 5. 死亡危機と作況との関係

筆者が構築を続けている「江戸時代における人口分析システム(DANJURO)[14][15][16]」を構成する寺院「過去帳」分析システムに、A寺とB寺の過去帳を登録して、基礎的な死亡分析を行った。

A寺とB寺の寺院「過去帳」に供養されている死亡者は、1760年から1799年に至る40年間で年間平均31.6人、死亡性比は119.9である。童子、童男、童女などの戒名の位号を持つ年少死亡者は、年間平均6.6人、死亡性比は161.4であり、死亡総数の19.4%を占める(図2)。年間死亡数が1760年から1799年に至る40年間の平均死亡数の2倍を超えたのは、1784

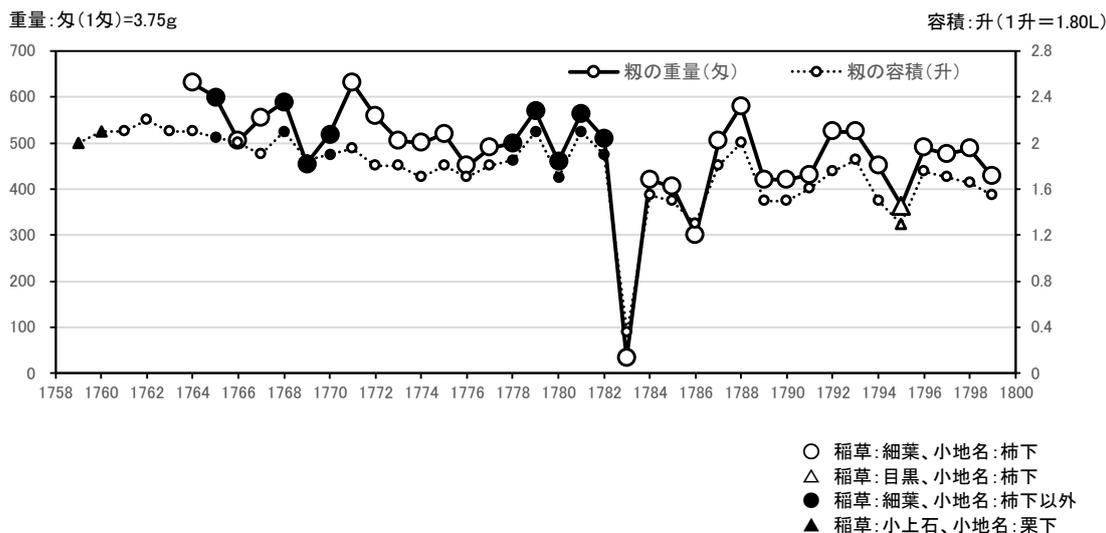


図1 陸奥国会津郡金井沢村における上田1坪から収穫された粃の収量(1759-1799)

(史料) 奥会津博物館架蔵、「作毛位付帳」(室井家文書、4196)より作成。

年の92人、1768年と1780年の64人であり、1776年の50人がこの3年に次いで死亡数が多い。

18世紀中期から20世紀初頭に至る期間で、1坪から収穫される籾の収量が最低を記録した1783年の死亡数は39人、死亡性比は171.4、翌1784年の死亡数は92人、死亡性比は152.8となった。18-19世紀の200年間にA寺とB寺の寺院「過去帳」に供養されている年間死亡数が、1784年に最多となり、男性死亡数が女性死亡数を大きく上回った。高野組では、1784年に18-19世紀最大の死亡危機を迎えた。

天明三年と天明六年の凶作を挟む安永十年(1780)から天明八年(1788)の「人数増減差引之覚(人数増

減改)」(室井家文書、1624、1626、1627、1629、1632、1633、1634、1635、1638)には、前年の「宗門改人別家別帳」が作成されてからの出生者、死亡者、転入者、転出者の異動内容が一人一人記録されている。1784年2/3月から1785年3/4月までの金井沢村における人口増加率は-76.0%、社会増加率は-55.6%、転入率は26.3%、転出率は81.9%、自然増加率は-20.5%、粗出生率は14.6%、粗死亡率は35.1%である。粗死亡率は、1780年から1781年3月までの期間に次ぐ高率となった。1784年2/3月の人口の4.5%が、挙家離村して行方不明となった。月間死亡数は1783年11月

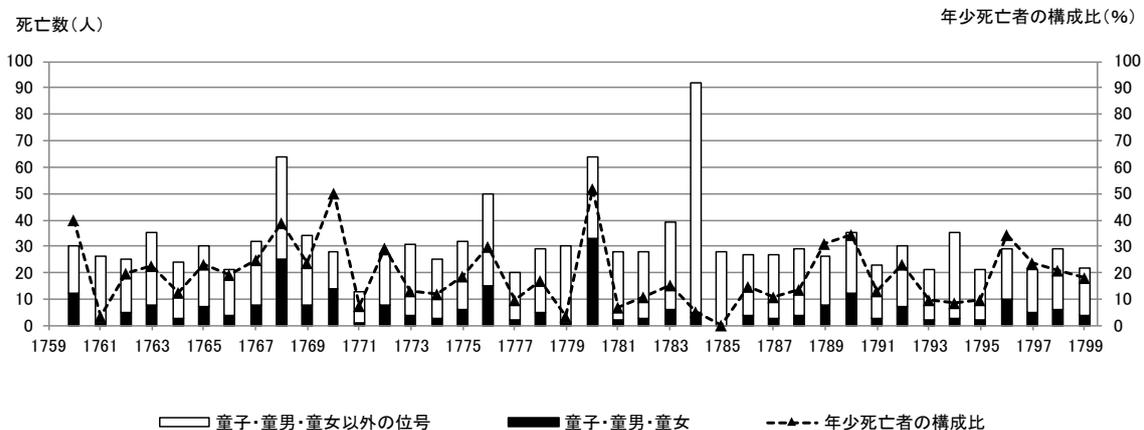


図2 陸奥国会津郡高野組の寺院「過去帳」に供養されている戒名の位号別死亡数(1760-1799)「江戸時代における人口分析システム(DANJURO)」の寺院「過去帳」分析システムにより作成。

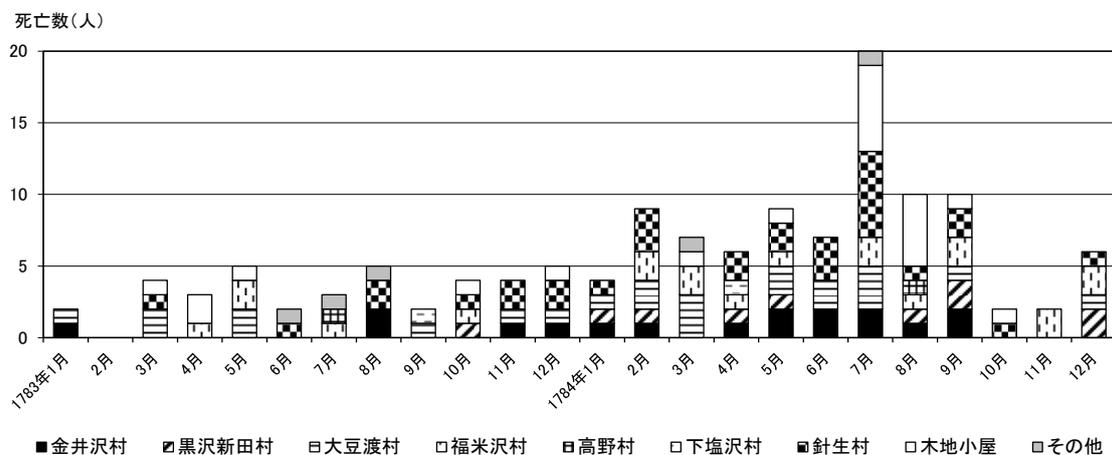


図3 陸奥国会津郡高野組の寺院「過去帳」に供養されている月間死亡数(1783-1784)「江戸時代における人口分析システム(DANJURO)」の寺院「過去帳」分析システムにより作成。

から 1784 年 9 月の期間に分散しているため、感染症が流行した可能性は低い(図 3)。

一方、天明 6 年(1786)に上田 1 坪から収穫した粃の重量は 300 匁、容積は 1 升 3 合であり、1759 年から 1799 年に至る期間で天明三年に次ぐ凶作となった(図 1)。1786 年秋にも、若松から検見役人が派遣されて、年貢の減免措置が取られた。蕎麦、粟、稗の作況も下と判定されている。しかし、1786 年と 1787 年の死亡数は、いずれも 27 人であり、40 年間の平均年間死亡数 31.6 人を下回った(図 2)。凶作のために年貢が減免された年であっても、凶作の程度により、死亡数は大きく異なった。

図 1 によれば、1768 年、1776 年、1780 年は、上田 1 坪から収穫された粃の重量が 450 匁、容積が 1 升 7 合を超え、平年を上回る作況であった。そのため、この 3 年は稲の作況と無関係な死亡危機とみられる。死亡性比は 1768 年に 120.7、1776 年に 138.1、1780 年に 106.5 であり、40 年間の平均死亡性比 119.9 に比較的近い。童子・童男・童女といった成名の位号を持つ年少死亡者の年間死亡数に占める構成比は、1768 年に 39.1%、1776 年に 30.0%、1780 年に 51.6%であり、いずれも 40 年間の年少死亡者の構成比 19.4%を大幅に上回っている(図 2)。年少死亡者の性比は、1768 年が 127.3、1776 年が 114.2、1780 年が 88.9 であり、40 年間の年少死亡性比 161.4 を大幅に下回る。稲の作況と連動しない死亡危機年の死亡性比は比較的均衡しており、年少死亡者の構成比が高く、年少死亡者の性比も比較的均衡している。そのため、1768 年、1776 年、および 1780 年に、性別を問わず免疫を持っていない年少人口に犠牲者の多い天然痘や麻疹などの感染症が流行した可能性を指摘することができる。

18 世紀後期の高野組では、1768 年、1776 年、1780 年、1784 年に死亡数が平常年の 2 倍を超える死亡危機が発生した。このうち、冷害にともなう凶作と連動していたのは、天明 3(1783)年の凶作の翌 1784 年に確認できる 18-19 世紀最大の死亡危機だけであった。1784 年には、年少死亡者の構成比が極めて低く、死

亡性比が不均衡で男性の死亡者が多数を占めた。一方、平常年を上回る豊作であった 1768 年、1776 年、1780 年の死亡危機は、年少死亡者の構成比が高く、死亡性比と年少死亡者の死亡性比が比較的均衡していた点に特色がある。他方、天明 3 年に次ぐ凶作となった天明 6(1786)年と翌年の死亡数は、平常年を下回っており、死亡危機は確認できない。死亡危機の要因は、凶作だけではないことが史料から裏づけることができた[17]。

## 6. 古天気、作況、死亡データベースの構築に向けて

18 世紀以降の東北地方における死亡危機の主要因を冷害にともなう凶作に求める見解には、再検討の余地が多いことが、前節の検討によって確認された。そこで、18-19 世紀の天気、作況、死亡者を記録した史料を蓄積、分析、公開するデータベースを構築して、天気や作況が死亡に与えた影響を評価することにより、死亡危機の要因を探る研究方法を模索したい。

### (1) 古天気データベースの構築に向けて

江戸時代の天候記録に関わるデータベースを構築した先行研究として、「歴史天候データベース」がある[18][19][20][21]。全国 20 カ所以上の日記などから天候記録を抽出して、毎日の天気表と天気地図を画面に表示する操作性の優れた web システムである。しかし、原史料の天候に関わる記録内容が公開されていないため、原史料の天候記録と天気表や天気地図に記号で表示される天気との対応関係が、利用者には不明である。原史料に寒暖、風向、気象災害などに関する記録があれば、気候復原に活用する方法を利用者とともに検討することも課題とみられる。

「歴史天候データベース」には、「データベースのオリジナル・データファイルは公開準備中です。」と表示されている。そのため、公開予定の天候記録、および福島県伊達郡飯坂村渡部家に保存されている日記の天候記録などを登録できるように、データベースを設計したい。多様な史料の天候記録を登録するには、史

料によって特徴のある天候表現や語彙に関する理解を深める必要がある。

「農業日記」に記録されている毎日の天気や農作業をもとに、集落位置情報テーブル(データ項目:国郡村名、緯度、経度、標高)、天気情報テーブル(データ項目:国郡村名、年月日(和暦)、年月日(西暦)、天気、体感温度・地震・気象災害などの関連情報、農作業、古文書画像)、および史料書誌情報テーブル(データ項目:国郡村名、史料作成者名、史料表題、史料作成年(和暦)、史料作成年(西暦)、史料表紙画像)から構成される「古天気データベース」を構築したい。

「古天気データベース」から毎日の天気を表形式で表示する「古天気表示プログラム」と、毎日の天気を地図上に時系列アニメーション表示する「古天気・地図表示プログラム」を開発したい。さらに、歴史気候学の専門家とともに、「古天気データベース」に蓄積される天気情報にもとづいて、日単位の気温や日射量などの気候要素を復元したい。気候要素の復元には、平野らが提案した降水日数から暖候期の気温を推計する方法[22]、市野らが提案した天気を階級区分して全天日射量を推計する方法が[23]、本研究に適用できるか可能性があり、後述する作況と気候要素との関係をモデル化して、作況の欠損値を推計する際にも有効とみられる。

## (2) 作況データベースの構築に向けて

「作毛位付帳」に記されている1坪から収穫した粳の収量をもとに、集落位置情報テーブル(データ項目:国郡村名、緯度、経度、標高)、稲の作況情報テーブル(データ項目:国郡村名、歩刈の年月日(和暦)、歩刈の年月日(西暦)、歩刈の行われた小地名、田主、田の等級、稲の品種、株数、粳の容積、粳の重量、1升当たりの重量、古文書画像番号)、畑作物の作況情報テーブル(データ項目:国郡村名、年(和暦)、年(西暦)、煙草、麻、大豆、小豆、蕎麦、粟、稗、芋、大根)、および史料書誌情報テーブル(データ項目:国郡村名、史料作成者名、史料表題、史料作成年(和暦)、史料

作成年(西暦)、史料表紙画像番号)から構成される「作況データベース」を構築したい。

本データベースに登録できる可能性のある史料として、19世紀の下野国那須郡両郷村と越後国魚沼郡湯之谷村における稲の作況記録などがある。多様な史料の作況記録を登録するには、稲の品種、容積、重量の理解を深める必要がある。

「作況データベース」から1坪から収穫された粳の重量、容積、株数の時系列変化や、品種別の株数と重量、株数と容積、重量と容積の相関を求める基礎統計をグラフ表示する「作況分析プログラム」と、作況に関わる諸指標を地図に時系列アニメーション表示する「作況・地図表示プログラム」を開発したい。

歴史気候学の専門家とともに、(1)で復元する暖候期の気温や全天日射量などの気候要素と稲の作況との関係をモデル化して、史料が散逸している天保12(1841)年から明治5(1872)年、明治11(1878)年から明治15(1882)年の作況を推計したい。「農業日記」には、苗代づくり、田植、草取り、稲刈をはじめ稲作に関する詳細な記録がある。そのため、(1)で復元する稲の生育期間における日単位の気候要素と作況との関係をモデル化できる可能性が高いとみられる。

## (3) 寺院「過去帳」分析システムの拡充

既開発のDANJUROを構成する寺院「過去帳」古文書画像データベースに、福島県南会津郡と大沼郡の寺院「過去帳」を追加登録する。A寺とB寺に加えて、南会津町の2カ寺、只見町の1カ寺の寺院「過去帳」の写真撮影を終了して、入力作業中である。奥会津の寺院「過去帳」には、死亡者の居住した村や枝村が記録されているため、村別の死亡性比、村別の戒名の位号別死亡数、村別の年少死亡性比などの死亡指標を算出する寺院「過去帳」分析プログラムを拡充するとともに、死亡指標を電子地図にアニメーション表示する「死亡指標・地図表示プログラム」を開発したい。

## (4) 死亡危機の要因分析に向けて

(2)で幕末維新期の欠損値を補完した作況データと(3)で得られた死亡指標を活用して、気候や作況な

どが死亡に与える影響を評価することにより、冷害にもなう凶作が死亡危機の主要因であったか検討したい。現段階では、異常気象、農業技術、作況、貢租、救慌・備慌、流通、物価、食物消費、感染症を含む地域構成要素間の関係のなかで、死亡危機の要因を捉える図4のような枠組みを想定している。

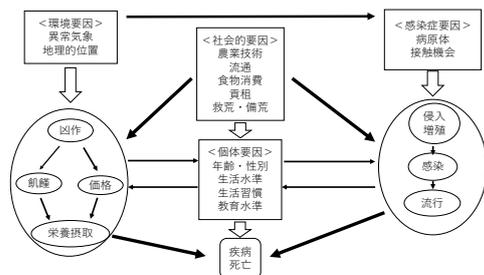


図4 死亡危機の発生メカニズム (鬼頭宏案を修正した)

## 7. おわりに

本稿では、18世紀初頭から100年以上におよぶ関東地方以北における人口減少の要因を解明するために、陸奥国会津郡金井沢村の名主家が、天気や農作業を記録した「農業日記」、1坪から収穫された粳の収量を記録した「作毛位付帳」、人口動態を記録した「人数増減差引之覚」、および死者を供養した寺院「過去帳」にもとづいて、18世紀後半の死亡危機と稲の作況との関係について展望した。検討の結果、凶作と死亡危機が連動していたのは、天明3(1783)年の大凶作と翌1784年の死亡危機にとどまり、平年作を上回る豊作年にも死亡危機が複数年にわたって確認された。

死亡危機の要因は、異常気象、農業経営、作況、貢租、救慌・備慌、食物流通、物価、食物消費、および感染症などを含む地域構成要素間の関係の中で捉える必要がある。そのためには、18-19世紀の天気、作況、死亡者を記録した史料を蓄積、分析、公開するデータベースを構築して、気候や作況などが死亡指標に与えた影響を評価することにより、死亡危機の要因を絞り込む研究方法を模索することが求められる。

## 参考文献

[1] 高橋梵仙, 日本人口史之研究第一, 三友社, 1941.  
 [2] 関山直太郎, 近世日本の人口構造, 吉川弘文館,

1958.  
 [3] 青木大輔, 寺院過去帳からみた岩手県の飢饉, 奥羽史談会, 1967.  
 [4] 菊地万雄, 日本の歴史災害 —江戸後期の寺院過去帳による実証—, 古今書院, 1980.  
 [5] 速水 融, 近世奥羽地方人口の史的序説, 三田学会雑誌, 1982, vol.75, no.3, pp.70-92.  
 [6] 成松佐恵子, 江戸時代の東北農村, 同文館出版, 1992.  
 [7] 高橋美由紀, 在郷町の歴史人口学 —近世における地域と地方都市の発展—, ミネルヴァ書房, 2005.  
 [8] 高木正朗編著, 18・19世紀の人口変動と地域・村・家族 —歴史人口学の課題と方法—, 古今書院, 2008.  
 [9] 川口 洋, 17~19世紀の会津・南山御藏入領における人口変動と出生制限, 歴史地理学, 1998, vol.40, no.5, pp.5-25.  
 [10] 木下太志, 近代化以前の日本の人口と家族, ミネルヴァ書房, 2002.  
 [11] 佐藤常雄, 日本稲作の展開と構造 —坪刈帳の史的分析—, 吉川弘文館, 1987, p.265.  
 [12] 森 嘉兵衛・谷川健一編, 日本庶民生活史料集成 第7巻, 三一書房, 1970, p.407.  
 [13] 田島町史編纂委員会, 田島町史 第6巻(上), 歴史春秋社, 1986, p.655.  
 [14] “江戸時代における人口分析システム (DANJURO)”, <http://www.danjuro.jp/>, (参照2020-01-24).  
 [15] Kawaguchi, Hirosh, Data Analysis System for Population and Family Studies on Japan in the 17th-19th Centuries, *Japanese Journal of Human Geography*, 2009, vol. 61, no.6, pp.2-22.  
 [16] 川口 洋, 一八・一九世紀を対象とした人口・家族研究のための情報システム —宗門改帳・過去帳・戸籍を入力史料として—, (落合恵美子編, 徳川日本の家族と地域性 —歴史人口学との対話—, ミネルヴァ書房, 2015), pp.446-492.  
 [17] 川口 洋, 天明三年(1783)の冷害にもなう人口変動, (井上 孝・和田光平編, 災害の人口学, 原書房, 近刊).  
 [18] “歴史天候データベース”, <http://jcdp.jp/historical-weather-database-jp/>, (参照2020-01-24).  
 [19] 吉村 稔, 古気候の復元と歴史天候データベース, 地学雑誌, no.102, 1993, pp.131-143.  
 [20] 吉村 稔, 歴史天候データベース オン・ザ・ウェブについて, 天気, no.54, 2007, pp.83-86.  
 [21] 吉村 稔, 古日記天候記録のデータベース化とその意義, 歴史地理学, 2013, vol.55, no.5, pp.53-68.  
 [22] 平野淳平・三上岳彦・財木真寿美, 広島の日記天候記録による1779年以降の夏季気温の復元, 地理学評論, 2018, vol.91, no.4, pp.311-327.  
 [23] 市野美夏・三上岳彦・増田耕一, 日記天候記録から推定した日本における19世紀前半の日射量変動, 地学雑誌, 2018, vol.127, no.4, pp.543-552.

# 精緻な表記情報を有する「延喜式祝詞」コーパスの構築

## Construction of the Corpus of “Engishiki Norito” with Detailed Annotation of Notation

間淵 洋子

MABUCHI, Yoko

国立国語研究所, 東京都立川市緑町 10-2

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

**概要:** 本論文では、現在開発を進めている「延喜式祝詞」のコーパスについて、その仕様と整備状況について報告する。延喜式祝詞は、上代から中古にかけての日本語を研究するための言語資源として期待される資料で、万葉仮名の使用や、宣命書きといった特徴的な表記形態が見られるほか、現存する写本には、ヲコト点や傍訓などの漢文訓読用の書き入れが施されている。本研究では、これらの表記特徴をできるだけ精緻に写したXMLによる構造化テキストを整備する。加えて、訓読文に対する形態素解析についても検討を進める。

**Abstract:** In this paper, we report on the specification and developmental status of the corpus of "Engishiki Norito" which is currently under construction. "Engishiki Norito" is considered a valuable material as a language resource for studying Japanese of Nara-Heian period because it has characteristic notation styles such as "manyogana", "senmyo-gaki", "wokoto-ten", "kana-ten", etc. We are developing XML structured texts with detailed annotation of these notational features. In addition, we consider morphological analysis of Sino-Japanese gloss reading sentences for the corpus with morphological information.

**キーワード:** コーパス, 延喜式祝詞, 表記情報, XML

**Keywords:** corpus, “Engishiki Norito”, annotation of notation, XML

### 1. はじめに

現在、国立国語研究所(以下「国語研」)では、上代から中古にかけての日本語を研究するための言語資源として、「延喜式祝詞」(『延喜式』巻八所収の祝詞)のコーパス構築を進めている。

「延喜式祝詞」は、内容語や活用語の活用語幹を通常の文字大で、助詞・助動詞などの付属語や活用語の活用語尾を小書きの万葉仮名で表記する、上代日本語の独特な表記法「宣命書」に表記上の特徴がある。また、現存する『延喜式』の写本には、本文筆写の後に、ヲコト点や仮名点(付訓)等、漢文訓読用の書き

入れ(訓点)が施されている点においても、日本語研究資料としての価値が高い。

昨今、典型的な和文とは性質を異にする宣命や訓点資料といった資料を電子化し、有用なアノテーションを加えることで、計量的にこれらの資料における語彙・語法や表記の実態を捉えるための試みが進められているが[1][2]、両資料の性質を併せ持つ「延喜式祝詞」のコーパス化が実現すれば、万葉仮名や宣命書の表記傾向、訓点の使用実態等を比較することにより、資料間の差異や個々の資料の特徴を浮き彫りにすることができよう。

そこで、本研究では、表記史研究に資する言語資源としてのコーパス形態のプロトタイプを策定することを目指し、精緻な表記情報を付与した「延喜式祝詞」のコーパス構築を試みる。本稿では、主に、(1)延喜式祝詞の資料特性、(2)表記情報を精緻に写すためのデータ仕様、(3)データ構築方法、の3点について報告する。

## 2. 延喜式祝詞

『延喜式』は、平安時代中期に律令の施行細則をまとめた法典(「格式」)で、延喜5(905)年8月、醍醐天皇より命を受けた藤原時平により編纂が開始された。延喜に撰修を始めたことに、『延喜式』の名前の由来があるが、その完成は22年後の延長5(927)年、施行は康保4(967)年である。全50巻中、巻一から巻十までが神祇祭祀に関する規定を収めたもので、そのうちの巻八が「祝詞」の巻となっている[3]。

「祝詞」は、神前に奏上する際の独特の文体を持つ言葉で、『延喜式』巻八所収の祝詞(以下「延喜式祝詞」)は、現存する最古の祝詞とされる。「延喜式祝詞」は27の祝詞を所載する。それぞれの祝詞の成立は、祭祀の開始時期により奈良朝以前から平安初期までとされているが、「延喜式祝詞」自体の作成時期はその時期と一致しないという指摘もあり[4]、安易に上代日本語資料として扱うことには問題があるものの、残存する資料の少ない上代・中古初期において貴重な資料の一つである。

祝詞には、体言や用言の語幹を通常の文字大で、助詞・助動詞や活用語尾などを万葉仮名の小書き右寄せ(または2行の割書)で表記する独特の表記法が用いられている。この表記法は、祝詞と同様、このスタイルで記された「宣命」(天皇の勅命を和文で記した文書)にちなみ「宣命書」と呼ばれている。

『延喜式』の原本は現存しないが、伝本は少なくない。そのうち、巻八を存するもので代表的な古写本には、「九條家本」「卜部兼永自筆本」「卜部兼右自筆本」などがあるが、中でも「九條家本」(東京国立博物

館蔵)は、全50巻中27巻を存し、平安中期の書写と見られる現存最古の写本として貴重である。現在、国立文化財機構が提供するWebコンテンツ「e国宝」(<http://www.emuseum.jp/>)により高精細画像を閲覧することができる。



図1 「延喜式 巻八(九條家本)」1巻  
(東京国立博物館蔵, e 国宝)[5]

九條家本巻八には、墨点による仮名点(付訓)やマコト点(マコト)が施されており、平安中後期加點資料として貴重なものである。

そこで、本研究では、最古の写本、九條家本を底本とし、緻密な翻刻とこれに基づく訓読文を提示した『東京国立博物館蔵本 延喜式祝詞総索引』[3]に基づき、九條家本「延喜式祝詞」のコーパスを開発することとした。その際、万葉仮名、宣命書、訓点などの特徴的な表記情報を精緻に写し取り、日本語表記史研究に資するコーパスとすることを目指した。

## 3. 電子化フォーマット的设计

コーパス構築にあたっては、まず、資料を電子化するための仕様を定める必要がある。本研究において

は、「延喜式祝詞」の資料性および研究目的に鑑み、設計方針を以下の通り設定した。

- (1) 記された文字をできるだけ忠実に写す。
- (2) 宣命書や訓点等の表記情報を明確に表現する。
- (3) 文書要素・言語単位と表記との対応関係を明確に示す。
- (4) 原文(翻刻本文)と訓読文との対応関係を明確に示す。
- (5) 原文と訓読文をそれぞれ抽出できるようにする。
- (6) 分析・解析用ツールで扱いやすい形とする。
- (7) 情報の追加を想定し、拡張しやすい形とする。
- (8) 底本と容易に対照できるようにする。

この設計方針に基づき、データの電子化フォーマットを以下の通り定めた。

- 文字コードは UTF-8, 文字セットは JIS X0208 に準じる。
- XML 形式による構造化テキストとする。以下のタグにより情報を格納する。

#### 【文書構造情報を表すタグ】

- text** テキスト全体(延喜式における1巻)。属性に書誌情報を含む。  
 @title="延喜式"  
 @volume="巻八"  
 @source="九條家本"
- title** 文書または特定の文章範囲の見出し。
- div** 文章範囲(各祝詞の範囲)。属性に各祝詞の情報を含む。  
 @title="祈念祭(例)"  
 @type="(奏上|宣命)"
- s** 文
- lb** 原典における行頭位置。属性に行番号(@n)を含む。『東京国立博物館蔵本 延喜式祝詞総索引』において認定・付与された行番号に一致する。

#### 【表記情報を表すタグ】

- span** 特殊な表記様式を持つ範囲。属性に、種別(type)を持つ。主に割注(割書)箇所を示す。
- ruby** 傍訓とその対象本文の対。
- rt** 本文に傍書されるテキスト(傍訓)。属性に、原文文字列(@originalText;任意)を持つ。
- rb** 傍訓の対象となる本文行のテキスト。熟字訓を除き1文字を単位とする。
- corr** 原文(翻刻本文)からの修正(補読等)。属性に校訂情報を含む。  
 @type="(omission|excess|erratum|hendokuB|hendokuA)" 脱字 or 衍字 or 誤字 or 返読前 or 返読後  
 @resp="(editor|annotator)" 校注者 or 作業者  
 @id(任意, 返読前後を対応付ける固有番号)
- kunten** フコト点・仮名点等の訓点による訓読。属性に文字情報を含む。  
 @type="(wokoto|kana)" フコト点 or 仮名点  
 @originalText(任意, 仮名点の場合のみ)
- kana** 大書きの万葉仮名, 宣命書による文字。  
 @type="(manyo|senmyo)" 万葉仮名 or 宣命書  
 @originalText(万葉仮名字体)

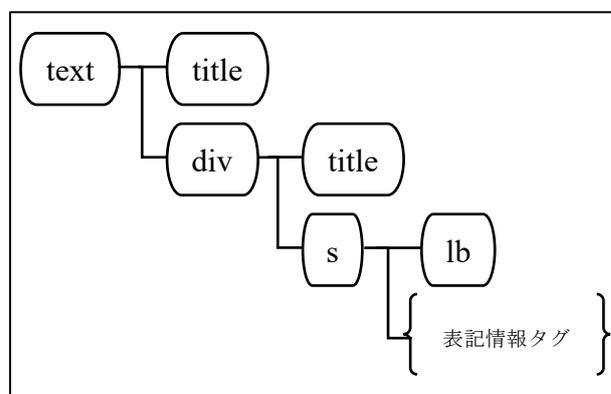


図2 文書構造情報に関する要素のスキーマ

## 4. データ構築作業

データ構築にあたっては、コーパスの底本を『東京国立博物館蔵本 延喜式祝詞総索引』所収の訓読文と定め、これを文字入力した。次に、入力本文と「翻刻本文」「影印」(原典画像)とを対照しながら、訓読文との差異を3節に示したタグを用いて写し取った。作業の効率化を図るため、簡易的なタグを使用しタグ付けしたものを一括で XML タグへ変換する方法で実施し

た。「延喜式祝詞」に出現する特徴的な表現と、その電子化形式について、以下に例示する(用例画像は、いずれも『東京国立博物館蔵本 延喜式祝詞総索引』より引用した)。

【宣命書と万葉仮名】

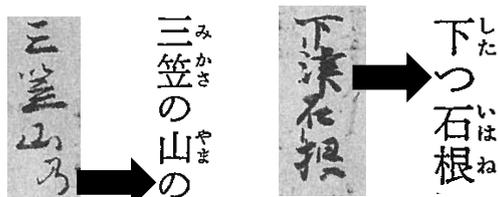


図3 原文・訓読文の対応(左:宣命書, 右:万葉仮名)

```
<ruby><rb>三</rb><rt>み</rt></ruby>
<ruby><rb>笠</rb><rt>かさ</rt></ruby>
<corr type="omission" resp="editor">の</corr>
<ruby><rb>山</rb><rt>やま</rt></ruby>
<kana type="senmyo" originalText="乃">の</kana>
```

図4 宣命書の形式化

```
<ruby><rb>下</rb><rt>した</rt></ruby>
<kana type="manyo" originalText="津">つ</kana>
<ruby><rb>石</rb><rt>いは</rt></ruby>
<ruby><rb>根</rb><rt>ね</rt></ruby>
```

図5 万葉仮名の形式化

【訓点】

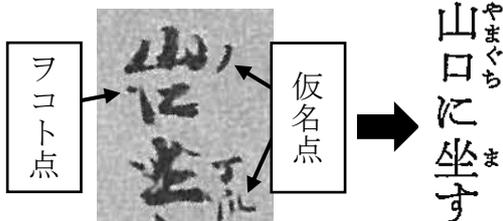


図6 原文・訓読文の対応(ヲコト点と仮名点)

```
<ruby><rb>山</rb><rt>やま</rt></ruby>
<corr type="excess" resp="editor"><kunten
type="kanaten" originalText="/" /></corr>
<ruby><rb>口</rb><rt>ぐち</rt></ruby>
<kunten type="wokoto">>こ</kunten>
<ruby><rb>坐</rb><rt originalText="マ">ま
</rt></ruby>
<kunten type="kanaten" originalText="ス">す
</kunten>
```

図7 訓点の形式化

【漢文式語順】

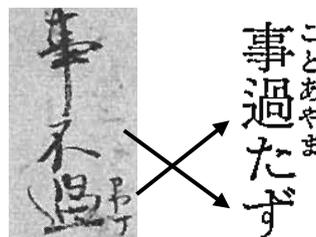


図7 原文・訓読文の対応(漢文式語順)

```
<ruby><rb>事</rb><rt>こと</rt></ruby>
<corr type="hendokuB" resp="editor"
id="07401">不</corr>
<ruby><rb>過</rb><rt originalText="アヤマ">
あやま</rt></ruby>
<corr type="omission" resp="editor">た</corr>
<corr type="hendokuA" resp="editor"
id="07401" originalText="不">ず</corr>
```

図9 漢文式語順の形式化

上記 XML による構造化が終了した後、原文(翻刻本文)、訓読文の各テキストを抽出するテストと、タグ付けのチェックとを兼ねて、XSL により、訓読文・原文を HTML 文書へと変換、出力し確認した。その際、要素や属性によって表示を工夫することで、タグ付けの実

態を容易に判別でき、底本となる訓読文、あるいは原文との照合がしやすい環境を整えた。



図10 訓読文 XHTML の Web ブラウザ表示



図3 翻刻本文 XHTML の Web ブラウザ表示

これにより、原文との関係性を明確にしつつ、原典の表記情報を精緻に写し取った「延喜式祝詞」の訓読文ベース XML データが完成した。

### 5. 今後の課題と展開

本研究では、日本語表記史研究に資する言語資源としてのコーパス構築を目指している。ここまでの整備状況として、前節までで述べてきた通り、表記情報を精緻に写し取った訓読文ベースの本文を XML 形式により作成したが、今後、研究利用に供するためには、少なくとも以下の二つのフェーズで開発・検討を進める必要があると考える。

まず一つは、形態論情報の付与である。

日本語表記史研究においては、どのような言語要素、どのような語に対して、どのような表記を用いているかといった、表記と言語単位との相関の解明が大きな課題の一つであるが、計量的手法によりこれを可能にするためには、適切な言語単位に分割しておくことが不可欠となる。すなわち、形態素解析を施す必要がある。

日本語の古代語・近代語に対する形態素解析手法については、国語研が開発を進める「日本語歴史コーパス」([https://pj.ninjal.ac.jp/corpus\\_center/chj/](https://pj.ninjal.ac.jp/corpus_center/chj/))の構築過程において、これまで多くの試みがなされており、各時代の日本語資料に即した解析用辞書の開発が行われてきた[6][7][8]。しかし、それらはあくまでも漢字仮名で表記された和文を対象としたものであり、漢字文に適用できるものではない。そのため、本研究でも、ベースとなる本文を原文ではなく漢字ひらがな表記による訓読文として整備を進めてきた。これにより、既存の形態素解析システムに適用可能な電子データの準備は整ったが、「延喜式祝詞」は表記のみでなく、語彙の面でも既存の上代・中古の作品とは、大きく異なっている。現在、上代(万葉集用)と中古(和文)の2種の解析辞書を用いて形態素解析試行を実施しているが、いずれも解析精度は高くない。今後、3月に国語研から「日本語歴史コーパス 奈良時代編II 宣命」として公開予定の五国史宣命を資料としたコーパスなども用いて、「延喜式祝詞」の本文解析に適した辞書を開発することも、残された課題の一つである。

なお、形態論情報アノテーションの結果は、国語研の「日本語歴史コーパス」との連携を確保し、国語研のコーパス検索用オンラインシステム「中納言」での公開を視野に開発を進める予定であり、言語単位として、現在国語研で公開されているコーパスが統一的に採用している「短単位」[9][10]を使用する。その際、上代語、あるいは祝詞や宣命に特有の問題点として、語形(語の読み方とそれに伴う単語の区切り方)と表記単位の不整合(「王臣等」の文字列を「おほきみたちまへつきみたち(=大君/達/公卿/達)」と読む類)が生じるため、これらの対応についても検討が必要である。

さらに、今一つの課題として、校異情報の付与を挙げておきたい。

古体を保持する祝詞の性質上、「延喜式祝詞」は本研究で底本とした九条家本に代表される古写本の他にも、長く後世に伝わっており、伝本が多数存在する。本研究プロジェクトの連携機関でもある歴史民族博物館(以下「歴博」)が所蔵する、江戸初期の写本「土御

門家旧蔵本」もその一つであり、金子(2012)により巻八祝詞の翻刻も試みられている[11]。史料画像の共有によって加点情報などを含めた異同を確認することができる。できれば、筆写時期による表記・形態の差異を明らかにすることができる。その他、「延喜式祝詞」諸本の詳細な校異を有する金子(2014) [12]など、「延喜式祝詞」の異文研究の成果を統合的に格納したコーパスというのも、本研究で目指すべき一つの形態であると考えられる。

なお、校異情報の付与は、TEI (Text Encoding Initiative; <https://tei-c.org>) による電子化フォーマットが確立されており、これに準拠した形式でのデータ構築が有用であると思われる。それと同様に、祝詞特有の表記体系をアノテーションするために本研究で策定した枠組みも、より標準的な枠組みとの連携を視野に検討を続ける必要があるだろう。

## 6. おわりに

本稿では、現在構築を進めている「延喜式祝詞」のコーパスについて、以下を報告した。

- (1) 日本語の表記史研究に資するコーパスとして、祝詞の特徴的な文体・表記(宣命書、万葉仮名)や、古写本に見られる訓点の情報を、精緻に付与した XML 文書としてデータ構築を進めている。
- (2) 形態素解析処理を前提に、訓読文をベースとしたコーパス本文を整備した。訓読文と原文との対応を保持し、表記情報と単語情報の相互参照が可能な形式での公開を検討している。
- (3) 更なるアノテーションの拡張として、異本と底本(九条家本)との校異情報の付与を計画している。

## 付記

本研究は、人間文化研究機構広領域連携型基幹研究プロジェクト「異分野融合による「総合書物学」の構築」国語研ユニット「表記情報と書誌形態情報を加えた日本語歴史コーパスの精緻化」及び国立国語研究所プロジェクト「通時コーパスの構築と日本語史研究の新展開」による成果の一部である。

## 参考文献

- [1] 池田幸恵・須永哲矢, 「五国史」宣命コーパスの設計とその利用, 訓点語と訓点資料, 2015, no. 134, pp.98-80.
- [2] 柳原恵津子, 『金光明最勝王経』平安初期点の形態素解析用本文作成: その方法と問題点, 「東洋学へのコンピュータ利用」第31回研究セミナー発表論文集, 2019.
- [3] 沖森卓也編, 東京国立博物館蔵本 延喜式祝詞総索引, 古典研究会, 1995.
- [4] 沖森卓也, 日本古代の表記と文体, 吉川弘文館, 2000.
- [5] <http://www.emuseum.jp/detail/100162>
- [6] 小木曾智信, 旧仮名遣いの口語文を対象とした形態素解析辞書, じんもんこん 2012 論文集, 2012, pp.25-32.
- [7] 小木曾智信・小町守・松本 裕治, 歴史的日本語資料を対象とした形態素解析, 自然言語処理, 2013, Vol.20, No.5, pp.727-748.
- [8] 小木曾智信・市村太郎・鴻野知暁, 近世口語資料の形態素解析の試み, 第4回コーパス日本語学ワークショップ予稿集, 2013, pp.145-150.
- [9] 伝康晴・小木曾智信・小椋秀樹・山田篤・峯松信明・内元清貴・小磯花絵, コーパス日本語学のための言語資源: 形態素解析用電子化辞書の開発とその応用, 日本語科学, 2007, Vol.22, pp.101-123.
- [10] 小椋秀樹・小磯花絵・富士池優美・宮内佐夜香・小西光・原裕, 『現代日本語書き言葉均衡コーパス』形態論情報規程集 第4版(下), 文部科学省科学研究費特定領域研究「日本語コーパス」データ班, 2011.
- [11] 金子善光, 翻刻・歴史民俗博物館所蔵「延喜式 巻八祝詞」, 神社と実務, 2012, no.11, pp.95-74.
- [12] 金子善光, 翻刻・京都大学図書館蔵「陽明文庫本 延喜式・巻八・祝詞, 『文化史史料考證』刊行委員改編, 嵐義人先生古稀記念論集 文化史史料考證, 2014, pp.23-60.

## 第 25 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

---

2020 年 2 月 29 日発行

編集・発行 第 25 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 〒112-0012 東京都文京区大塚 3-1-1  
TRC-ADEAC 株式会社

---

ISSN 2188-1529



