

[研究ノート]

# 昭和女子大学近代文庫所蔵「與謝野晶子書簡」 コーパス化と基本設計<sup>1</sup>

—日本文コースプロジェクト5<sup>2</sup>の取り組みと今後の公開に向けて—

宮 寄 由 美

The Nichi-bun Course Project Efforts and Aim to Open to the Public:  
The Basic Design of the Corpus of Letters by Akiko Yosano  
by Showa Women's University Modern Library

MIYAZAKI Yumi

## Abstract

In this paper, we report on the basic design of the corpus of "Akiko Yosano's letters which are owned by the Modern Library of Showa Women's University". We have been working on this corpus as a part of course project in the Department of Japanese Language and Literature since the academic year 2022. These ninety-one letters were written by Yosano in her later years and addressed to Yosano's disciple Mitsuko Omi, and her family. In creating this corpus, we referenced the digital archive of these letters and the reprinted version published by Sugimoto and Otsuka (1991). In the future, we will collaborate with the digital archives in opening this corpus to the public.

*Key words:* Akiko Yosano's Letters (與謝野晶子書簡), Digital Archives (デジタルアーカイブ), Corpus (コーパス), Showa Women's University Modern Library (昭和女子大学近代文庫)

## 1. はじめに

本稿では、昭和女子大学（以下、本学）図書館内近代文庫に所蔵されている昭和女子大学與謝野晶子書簡集（以下、「與謝野晶子書簡」（本書簡））を、人間文化学部日本語日文学科（以下、本学科）日本文コースプロジェクトの一環として、2022年度よりコーパス<sup>3</sup>化を行ったその基本設計と入力規則について報告する。本書簡は、與謝野晶子晩年の愛弟子であった近江満子とその家族に宛てた91通である。コーパス化にあたっては、図書館所蔵の本書簡デジタルアーカイブの画像と杉本・大塚（1991）の翻刻版を参照した。本コーパスは今後、デジタルアーカイブとの連携を目指すものである。

## 2. 本学「近代文庫」と「與謝野晶子書簡」所蔵の経緯と内訳

### 2.1 本学「近代文庫」と「與謝野晶子書簡」所蔵の経緯

本学では、本学創立者である人見圓吉（筆名、東明）と学内共同研究の成果である『近代文学研究叢書』（1956年創刊開始）が1958年に第6回菊池寛賞を受賞したことを記念し、「近代文庫」<sup>4</sup>を設立した（昭和女子大学図書館：2019）。本書簡は、晶子晩年の愛弟子であった近江満子のご遺族から1953年

に寄贈された。その後、本学図書館内近代文庫に所蔵され、現在は貴重図書としてデジタルアーカイブとともに所蔵されている。

## 2. 2 「與謝野晶子書簡」の内訳

本書簡は、晶子が51歳から61歳（享年63歳）であった1929年（昭和4年）1月から1939年（昭和14年）2月までの10年間に書かれたものである。その期間には、『冬柏』の創刊や『新新訳源氏物語』の刊行、晶子自身の狭心症の発作や夫鉄幹の死などがあった。その間に投函、もしくは使いに持参させた全91通である<sup>5</sup>。

本書簡の宛先は、表1に示すように、主に近江満子に宛てたものであるが、満子との出会いのきっかけとなった満子の夫で医師であった近江湖雄三、そしてその家族に宛てたものも含まれる。本書簡の形式の内訳と行数（本稿では文数に相当する）について表2に示す。

表1：宛先		表2：書簡の形式	
	(行)		(行)
宛		書簡形式	
近江満子	770	絵葉書	539
近江湖雄三・近江満子	270	封書	475
近江湖雄三	107	葉書	111
近江渉	9	巻紙	33
近江亘	9	書簡図絵	11
近江もゝ子	8	便箋	11
近江シゲヤス	7		
総計	1,180	総計	1,180

## 3. コーパス化の背景

本章では、杉本・大塚（1991）の翻刻版の刊行を中心に、学内外への発信から、2000年代以降の学界の動向とともに、本書簡の資料性を活かすべく、日文コースプロジェクト5におけるコーパス化の取り組みの経緯を述べる。

### 3. 1 本学での取り組み

1953年に寄贈された「與謝野晶子書簡」は、1991年当時まで学外に未発表であった。その書簡を本学文学部教授であった杉本邦子と同短期大学部助教授であった大塚豊子両氏が翻刻し、『與謝野晶子未発表書簡』（以下、翻刻版）を出版した。さらに本学図書館では2023年現在、本書簡のデジタルアーカイブ化が行われている。

### 3. 2 コーパスとデジタルアーカイブ

2010年代以降、日本でも書き言葉、話し言葉、メールやソーシャルネットワークワーキングサービスの言葉などのコーパス化とそれを元にした研究が本格的に行われるようになった。さらに近年では各図書館のデジタルアーカイブをより広く研究に活用する試みとして、一部デジタルアーカイブのコーパ

ス化が行われ始めている。

### 3. 3 日本語日本文学科と日文コースプロジェクト5の取り組み

2020年代に入り「データサイエンス」の理解と実践が今まで以上に必要とされるなか、いわゆる文系とよばれる本学科においても、これまでビッグデータをデータマイニングするその手法と活用を学修する授業を展開してきた。

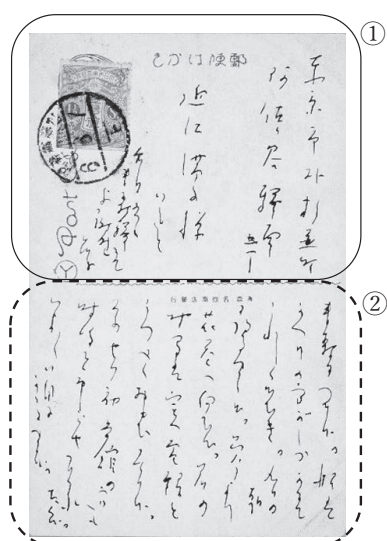
さらに筆者が着任した2022年度からは、そのデータマイニングの元となるデータベース（ここではコーパス）がどのような仕組みを持ち、どのように作成されているのかを学生との協働で解明すべく、「日文コースプロジェクト5」（筆者指導）にて、貴重書を多く所蔵する本学近代文庫の蔵書をコーパス化する活動を開始した。今回コーパス化するにあたり、近代文庫内貴重書の中でも、既に本学教員により詳細な翻刻版（杉本・大塚：1991）が作成され、図書館においてデジタルアーカイブ化もなされている点が本書簡を選定した根拠である。なお、本報告は2022年度末時点<sup>6</sup>での学生とともに行ったコーパス化作業のデータを元に行っている。

## 4. 本コーパスの基本設計

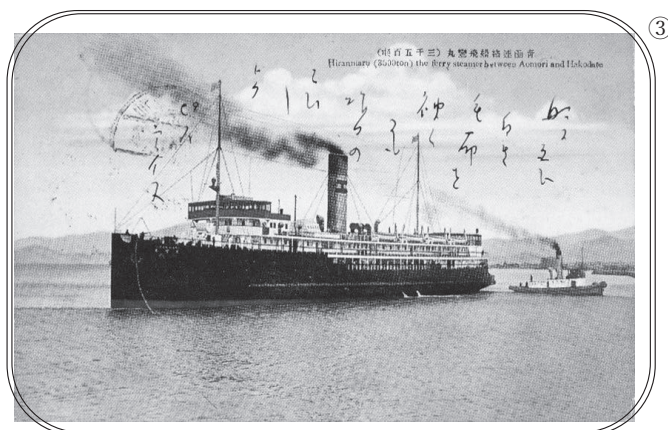
### 4. 1 書簡データと翻刻一本文記述箇所を再現する試み

画像1、2に示すのは、晶子から満子宛に青森県より1931年（昭和6年6月7日消印）に郵送された絵はがきである。

2. 2の表2の通り、本書簡の形式は「絵葉書、封書、葉書、巻紙、書簡図絵、便箋（原稿用紙含む）」である。これらには、例えば封書であれば封筒、葉書であれば宛先記載面、絵葉書の場合はさらに絵の掲載面にメッセージが記述されている場合などがある。そこで本コーパスでは、それらメッセージ記述箇所の再現のため、コーパス内に階層マークを付与することとした。この階層マークは、宛名書きがされている面を入力出発点とし、次の「メッセージ記述面の階層マーク」に示すようなメッセージ記述「面」1から3を設定し、コーパスの入力規則に反映させた。



画像1：晶子発満子宛「絵はがき」、面1（実線①）・面2（点線②） —管理番号 0020—



画像2：同「絵はがき」面3（二重線③） —管理番号 0020—

—メッセージ記述面の階層マーク—

面1：翻刻版において「表<sup>7</sup>」とされた部分。主に宛先記載面（例：画像1実線部①）

面2：翻刻版において「裏」とされた部分。封書などでは中身の本文面。（例：画像1点線部②）

面3：上記以外の面。もしくは、左記を本文とした場合の欄外に記載があった部分。（例：画像2二重線部③）

これらの規則を取り入れコーパス化したxlsxデータの具体例を、表3-1、3-2に示す。なお、ここ  
 という管理番号とはコーパス内での書簡の通番<sup>8</sup>を指す。

画像1の面1（実線①）、2（点線②）、画像2の面3（二重線③）のメッセージをxlsx形式で入力した  
 例が、それぞれ表3-1の実線①、表3-2の点線②、二重線③で囲んだ部分に該当する。詳しい入力規  
 則については4. 2「入力規則」の表4で述べることとする。

表3-1：晶子発満子宛「絵はがき」、面1（実線①）のコーパス該当部分 —管理番号0020—

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	管理 番号 A	管理 番号 B	宛	元表 記	日 付	消 印 発 行 所	消 印 日 付	消 印 時 刻	ジ ャ ン ル	形 式	面 1 自 筆 者	面 1 内 容	行 1	面 1 本 文
1														
228	0020	1	近江満子		昭和6年6月7日	□□青森	6・6・7	□□ - □	書簡	絵葉書	晶子	宛名	1	東京市外杉 並町阿佐ヶ 谷駅南二丁
229	0020	2	近江満子		昭和6年6月7日	□□青森	6・6・7	□□ - □	書簡	絵葉書	晶子	宛名	2	近江満子様
230	0020	3	近江満子		昭和6年6月7日	□□青森	6・6・7	□□ - □	書簡	絵葉書	晶子	宛名	3	御もと
231	0020	4	近江満子		昭和6年6月7日	□□青森	6・6・7	□□ - □	書簡	絵葉書	晶子	宛名	4	六月六日
232	0020	5	近江満子		昭和6年6月7日	□□青森	6・6・7	□□ - □	書簡	絵葉書	晶子	宛名	5	青森駅にて
233	0020	6	近江満子		昭和6年6月7日	□□青森	6・6・7	□□ - □	書簡	絵葉書	晶子	差出人	6	與謝野晶子
234	0020	7	近江満子		昭和6年6月7日	□□青森	6・6・7	□□ - □	書簡	絵葉書				
235	0020	8	近江満子		昭和6年6月7日	□□青森	6・6・7	□□ - □	書簡	絵葉書				

表3-2：晶子発満子宛「絵はがき」、面2（点線②）・面3（二重線③）のコーパス該当部分

	O	P	Q	R	S	T	U	V	W
	面 2 自 筆 者	面 2 内 容	行 2	面 2 本 文	面 3 自 筆 者	面 3 内 容	行 3	面 3 本 文	注 記
晶子	本文	1	青森につき候。						
晶子	本文	2	船はかへりの方がしづかにてうれしく候ひ ま。						
晶子	本文	3	九日の朝に帰るべく候。						
晶子	本文	4	只今より花巻へ向ひ候。						
晶子	本文	5	着の時間は宅へ電報をうつべくおもひ居り 候。						
晶子	本文	6	そのせつ初音館の方へも時間をおしらせ下 されたく候。						
晶子	本文	7	いよいよ御目にかかるべく候。						■いよく■ ■かゝる■
晶子	短歌	1	船に乗り白き毛布を被くにも昨 日のこひしトラビスティス						

## 4. 2 コーパスの入力規則

次の表4に示すのが、本コーパスの基本的な入力規則である。コーパスはxlsx形式で作成し、句点やスペース、改行を1文の目安とし、1行に入力した。

A列からJ列までは書簡の基本情報（発送年月日、宛名、日付、消印、書簡の形式）を入力した。また先述の通り、本コーパスではメッセージ記述「面」を取り入れており<sup>9</sup>、K列からN列が「面1」、O列からR列が「面2」、S列からV列が「面3」のそれぞれ「自筆者、内容、出現行、本文」の情報となっている。

表4：入力規則

列	タグ情報	入力規則
A列	管理番号A	書簡番号：4桁で入力 発送時期順を通番とし、書簡1を0001 <sup>10</sup> 、書簡11は0011
B列	管理番号B	書簡内の本文内出現行ナンバー 例：書簡1の5行目→5 書簡1の18行目→18
C列	宛	宛名情報を入力。D列に示すように書簡、翻刻版に表記のゆれがあった場合は、検索時の利便性を鑑み、本列に統一した記述を記した。主に「近江湖雄三・近江満子」の場合を指す。
D列	元表記	「近江湖雄三・近江満子」「近江湖雄三・令夫人」「近江夫妻」の3パターン。C列では「近江湖雄三・近江満子」に統一してある。
E列	日付	翻刻版の各書簡冒頭にある日付を入力。数字は算用数字。 使い持参便の場合もあり、その場合は満子による記録か、翻刻者による推測を含むため、F列以降の「消印」情報とは別に本列を設けた。 ただし、注記の必要な場合はW列にその内容を入力。 例：[書簡0004の場合] E列 昭和5年4月9日 W列 使い持参便 日付は満子の覚え書きによる
F列	消印発行所	消印スタンプのある場合はその情報を入力。数字は算用数字。 ①一部読み取り不能の場合は、文字数分□を入力。例：□□登別 ②文字数も読み取り不能の場合は全角でハイフン「-」を入力。 文字判別は翻刻版に則る。
G列	消印日付	「消印発行所」と同様、スタンプの押印が一部判別不能の場合は文字数分□を入力。文字数も不明な場合は「-」を入力。文字判別は翻刻版に則る。
H列	消印時刻	同上。
I列	ジャンル	「書簡」に統一。
J列	形式	書簡の形状を入力。名称は翻刻版に則る。 形式：絵葉書、封書、葉書、巻紙、書簡図絵、便箋
K列	面1自筆者	当該面の筆者情報を入力。 一部鉄幹の書、満子の走り書き、印刷などが存在。 筆跡の判別は翻刻者による。
L列	面1内容	基本的には翻刻版に則り、主に送受信関連の情報を入力。 [差出人、宛名、住所、名前、日付]
M列	行1	面1内、文単位の出現順を入力。
N列	面1本文	面1の記載内容を入力。 ①宛名であっても本文とみなし、1文1セルで入力。 ②踊り字は形態素解析を鑑み、本文行では「をろ／＼」「かゝり」を「をろをろ」「かかり」と入力し（近藤：2021）、注記（W列）に本来の表記を■で囲み入力。 ③旧字などMicrosoft excelのUnicodeで入力できない場合は「=」（ゲタ）を入力し、注記にその旨を入力しておく。

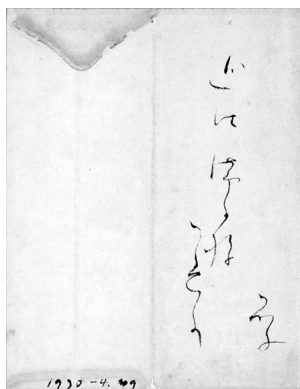


		④判読不能箇所は翻刻に則り、判別不能文字数分□を入力。
O 列	面 2 自筆者	K 列同様。
P 列	面 2 内容	基本的には翻刻版に則り、括弧 [ ] 内の内容を入力。 〔差出人、本文、短歌、詠み手〕 短歌、詠み手は本文の中でも特筆すべき内容と考え、今回カテゴリを設けた。
Q 列	行 2	面 2 内、文単位の出現順を入力。
R 列	面 2 本文	面 2 の記載内容を入力。入力規則は N 列と同様。
S 列	面 3 自筆者	K 列、O 列と同様。
T 列	面 3 内容	基本的には L・P 列と同様。 ただしこの面については、余白や裏面の走り書きなどが相当することがあり、次の例に示す情報が入力される場合もある <sup>11</sup> 。 例：翻刻版解説に「余白に…」とあれば余白、「裏面」とあれば裏面と入力。
U 列	行 3	面 3 内、文単位の出現順を入力。
V 列	面 3 本文	面 3 の記載内容を入力。
W 列	注記	本文特筆すべき点、気づいた点（疑問点とは別）があれば入力。また、本文中の踊り字を含む当該の形態素を■で囲みこちらに転記。 例：日付は満子による。 例：裏面の写真から線を引いて「ユフィン」「コノヘンガユフキン」と記している。 例：[R 列：あつくあつく] 本列：■あつく／＼■
X 列	【検討中】 画像	デジタルアーカイブの画像とのリンク。具体的には次章 4. 3、画像 3、4、表 5 で示すポップアップ画面を参照されたい。現在公開範囲とともにリンク利用許可申請中（2023 年 9 月）。

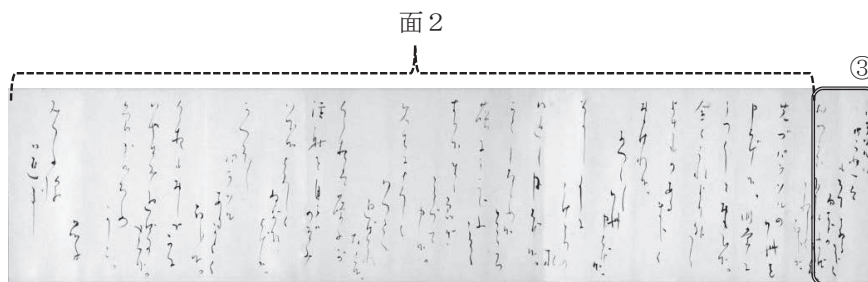
#### 4. 3 デジタルアーカイブとの連携を目指して

画像 3、4 は、1930 年（昭和 5 年）4 月 9 日に晶子から満子に宛てて送られた書簡（巻紙：管理番号 0004）である。本書簡には出版に関わる事務的な内容のものから、満子の家族に宛てたごくプライベートなやりとりまで記述されている。4. 1 で言及した通り、表記箇所もいわゆる「本文」とは別のマークを設けておきたい場合がある。例えば、画像 4 二重線③部のように、いわゆる本文とは別に、「冒頭余白」に走り書きとして「このパラソルははこねへはおちてまゐるまじく雨などのふらばあしく候。」と記されている例などがそれにあたる。これらを反映した文字列データ入力例を表 5 に挙げる。

管理番号 0004 とした書簡は、巻紙（画像 4）が封書（画像 3）に封入されている。そこで、階層マー



画像 3：「巻紙」面 1  
—管理番号 0004—



画像 4：同「巻紙」面 2、面 3（二重線部③） —管理番号 0004—

クとしては、画像3を面1、画像4の本文を面2として入力、最後に二重線で囲んだ部分（画像4の③部分）を面3として入力した。

この面3には、アノテーション情報として、「内容」列に「余白」との情報を付与してある。

これら階層性を反映させ入力した例が表5、51行目から57行目（画像4の後半部分に対応）と、画像5（画像4の冒頭余白部分③に対応。表5、58行目）である。このように、文字情報と画像とをリンクさせ、当該画像がポップアップする仕組みを構築することで、全体像や当該箇所の表記の把握が即時に可能となるよう作業を進めている。

表5：画像4 —管理番号0004— 面2、面3コーパス画面

	A	N	O	P	Q	R	S	T	U	V	W	X
1	管理番号 A	面1 本文	面2・1 自筆者	面2・1 内容	行2	面2・1 本文	面2・2 自筆者	面2・2 内容	行3	面2・2 本文	注記	【検討中】画像③
51	0004		晶子	本文	7	かへずがえすバラソルありがたく存じ候。					使用持参使 日付は満子の覚え書きによる	画像4_1.png
52	0004		晶子	本文	8	今夜もおしづかに御やすみ遊ばすべく候。					使用持参使 日付は満子の覚え書きによる	画像4_1.png
53	0004				9	急ぎ候ため					使用持参使 日付は満子の覚え書きによる	画像4_1.png
54	0004				10	かしこ。					使用持参使 日付は満子の覚え書きによる	画像4_1.png
55	0004				11	晶子					使用持参使 日付は満子の覚え書きによる	画像4_1.png
56	0004				12	みつ子様					使用持参使 日付は満子の覚え書きによる	画像4_1.png
57	0004				13	御返事					使用持参使 日付は満子の覚え書きによる	画像4_1.png
58	0004						晶子	余白	1	このバラソルははこねへはおちてまるまじく雨などのふらばあしく候。	使用持参使 日付は満子の覚え書きによる	画像4_1 冒頭余白.png

画像5（二重線部）：「巻紙」面3（画像4③部分） —管理番号0004—

## 5. おわりに

ここまで、学生が協働しその成果を学内外に発信する日文コースプロジェクトの一環として行った本学所蔵の「與謝野晶子書簡」のコーパス化と、2022年度時点での基本設計について報告した。2023年度には学生と試行し改善点の検討を試みている。最終的な設計とそれを元にした調査結果は別稿にあらためたいと考えている。

本学近代文庫には貴重書が多く収蔵されている。その中でも1953年に寄贈された「與謝野晶子書簡」は、寄贈から約40年の時を経て杉本、大塚両氏により書簡内容の背景とともに詳細な翻刻がなされた。それから約30年の時を経て、文系の学生にもデータサイエンスが研究の基盤となりつつある現在、貴重なデータを学生とコーパス化していくこのプロジェクトは、文学、言語学、自然言語処理学、図書館学など学際的な試みとして機能するものである。

本プロジェクトは2023年度も継続中であるが、言語学の視点からはまず、形態素解析の利便性を高めることを重視し、面2以降の情報をKからN列へ移行・集約していく。さらに4.3で述べたデジタルアーカイブとの連携作業を進め、より広く、様々な学術分野への貢献を目指していきたいと考える。

## 注

- 1 本稿は2023年8月29日に開催された「国立国語研究所言語資源ワークショップ（LWS）2023」にて発表を行ったポスターセッション原稿を元に、加筆修正を加えたものである。いずれも2022年度末までに整備したコーパスデータを元に報告を行うものである。
- 2 日本語日本文学科では2016年に「今とつながる」「地域とつながる」「世界とつながる」をキーワードにコースプロジェクトを発足した。日本の言語文化に関するテーマを元に、学生がそれぞれの専門知識を活かすことで協働し、成果を学内外に問う実践的な取り組みである。本稿でその成果を報告するプロジェクトはその5番目として2022年度に発足したプロジェクトである。
- 3 コーパスとは、言語資料を主にデジタルデータ化し、PCなどで検索可能なかたちにしたものである。
- 4 創立65周年を迎えた2023年現在も、図書館の重要な特殊文庫と位置付けられ、図書82,096冊、雑誌50,106タイトルを所蔵（昭和女子大学図書館：2023）。
- 5 主に晶子による直筆。一部與謝野鉄幹、満子による走り書き、短歌同人による筆、印刷も含まれる。
- 6 日文コースプロジェクト5の22年度活動時期は2022年10月から23年3月。日文の学生1年生から3年生（全18名）による作業結果である。参加者の氏名は本稿末に記す。コーパス化するにあたっては、①與謝野晶子や近江満子とその周辺の人物像や時代背景、②デジタルアーカイブ画像とのリンク作業を見据え、図書館との著作権、公開範囲の検討、③既存の、主に近代コーパスの設計に関する先行研究などを学生とともにまとめた。
- 7 杉本・大塚（1991）では、規則的に、書簡の推定も含む発信日、書簡の形式、消印が記述され、次に“表”として宛名と住所、“裏”として差出人と住所が翻刻されている。その後、本文内容が続き、最後に（余白）（裏面に）などとし、走り書き等が翻刻されている。後に示す画像4の太字で囲んだ部分などは、翻刻では本文の後に（余白に）とし、メッセージ内容の翻刻がある。
- 8 コーパス内での通番は、翻刻版の出現順（送付時期順）に則る。
- 9 2022年度の作業では、学生による入力段階で本文記載「面」の階層性を意識させる目的で0列以降の入力を行った。またそれは、起こり得る階層の判断に関するミスを修正しやすくする目的も兼ねている。23年度作業時には、あらたに「面情報」のような列を設け、KからN列に情報を集約していく予定である。
- 10 杉本・大塚（1991）の翻刻版には、近江家以外に送られた「晶子書簡」1通、「寛書簡」13通が収められている。今後コーパスへの追加も鑑み、暫定的に4桁とした。
- 11 本来規則的に、L・P列と同レベルの情報が入力されている必要があるため、「余白」「裏面」などの情報は今後別列を設ける必要がある。

## 参考文献

- 近藤明日子（2021）『コーパスと近代日本語書き言葉の一人称代名詞の研究』勉誠出版
- 昭和女子大学図書館（2019）『昭和女子大学図書館70年の歩み—図書館開設70周年・近代文庫創設60周年記念』デジタルパンフレット版 <https://swuhp.swu.ac.jp/library/zuroku-P/70th.pdf>
- 昭和女子大学図書館（2023）『2023年度図書館活動報告書』
- 杉本邦子・大塚豊子（1991）『昭和女子大学近代文庫所蔵 與謝野晶子未発表書簡』女性文化研究叢書第一集 昭和女子大学女性文化研究所

日文コースプロジェクト5 2022年度参加者（50音順）

秋山莉菜、浅野朱音、伊藤怜泉、大川桃未、折原悠理、國本蘭子、設楽優心、渋谷水美、清水美有、鈴木雅弓、鈴木結女、高松妃菜、竹下琴海、竜澤亜依、土田珠々、平田渚、藤原真希、村中美海

以上18名



## 謝辞

本コースプロジェクトの遂行や研究成果を執筆するにあたり、本学図書館長吉田昌志先生をはじめとする図書館の皆様には貴重な資料の利用・転載のご許可、さらには今後の公開に向けての検討を行っていただきました。また、本稿を掲載するにあたり、近代文化研究所所長の鳥谷知子先生や同研究所の皆様、査読にあたってくださった先生方のご尽力に感謝申し上げます。

(みやざき ゆみ 日本語日本文学科専任講師・近代文化研究所所員研究員)