

スタンフォード大学東アジア図書館所蔵「熊本藩文人書状集」

解読の試み

——第1 phase、第2 phaseを通して

マツザ 美恵子(スタンフォード大学東アジア図書館 テクニカルサービス部門長・
日本語部門テクニカルサービス司書)

E-mail miekom@stanford.edu

要旨

本稿は立命館大学アート・リサーチセンターの「日本文化資源デジタル・アーカイブ国際共同研究拠点 D. 研究設備・資源活用型」に筆者が2019年から参加した研究プロジェクトの第1 phase, 第2 phaseの活動記録である。スタンフォード大学東アジア図書館が所蔵する「熊本藩文人書状集」(以下、「書状集」)は購入時最小限のメタデータしか存在していなかった。当大学図書館オンラインカタログ([SearchWorks](#))での全文検索を可能にし、利用者の書誌情報へのアクセスを強化するため「AI くずし字解読支援・指導システム」を使用して解読作業を行ってきた。その第1 phase, 第2 phaseを通しての経験と今後の展望、解読されたコレクションの将来の可能性についてまとめた。

Abstract

Stanford University East Asia Library holds “Kumamoto-han bunjin shojoshu” (Collection of correspondence and poems from various officials from the Kumamoto domain) which is a collection consisting of letters and poems written in the cursive style and very difficult to read by a non-specialist. This article describes author’s experience using the ARC’s Kuzushiji AI database through phases 1 and 2 of the International Joint Digital Archiving Center for Japanese Art and Culture (ARC-iJAC), Art Research Center, Ritsumeikan University. Future planning of manuscript transcription projects is also discussed.

1. スタンフォード大学東アジア図書館の紹介

世界中でデジタルヒューマニティーズへの関心が高まり、デジタル画像データへのアクセスの需要が高まるにつれ、当大学図書館でも従来のMARCレコードによる書誌目録に加えて、画像資料をデジタル化し、デジタル画像相互運用のための国際規格 International Image Interoperability Framework (IIIF) に準拠した画像の掲載、全文検索などの機能が書誌情報に付加されることが主流になりはじめた。従来の書誌目録では、絵図の目録を作成する場合でも、米の図書館で使用されている Resource Description Access (RDA) 目録規則に従い、MARC フィールドの 500 フィールド(注記欄)に絵図に付記している情報を可能な限り記録するように努めていた。例えば、当館が所蔵している刷り物で「[乳姉妹, 善光寺如来靈驗記: 芝居刷りもの](#)」という資料に関しては、500 フィールドに日本

語とローマ字表記を併記して「明治四十二年三月一日午後三時開場菊池幽芳氏作乳姉妹七場春陽堂發行, 善光寺如来靈驗記三幕竹柴万二著作」と記録するのが精一杯であった(図1)。

500	6 880-03 a "Meiji yonjuninen sangatsu tsuitachi gogo sanji kajō Kikuchi Yūhō-shi saku u Chikyōdai nanaba Shunyōdō hakkō, Zenkōji Nyorai reigenki sanmaku Takeshiba Manji saku"
880	6 500-03 a "明治四十二年三月一日午後三時開場菊池幽芳氏作乳姉妹七場春陽堂發行, 善光寺如来靈驗記三幕竹柴万二著作"

図1 乳姉妹, 善光寺如来靈驗記: 芝居刷りもの. MARC 500 field.

また、全ての一枚摺の資料が順次デジタル化されていく訳ではないが、近い将来デジタル化される可能性を考慮して画像に付随している文章を[書き起こし](#)、500

フィールドに可能な限り記入するようにしていた(図2)

500 "忠臣義士四十七士の面々は二手に分れて表門よりは大石内蔵之助を総大将となし堀部弥兵衛以下忠義の面々得物々々を打振ひ表門玄関より乱入すれば寝耳に水の吉良の附人等は突然鼎の湧くが如く之に應對すると雖も忠義に凝りたる刃には刃向ふ術なく逃る者は追はず向ふ者は切捨てて室内に討ち入ったり"--Right side of image.

500 "Chūshin gishi shijūshichishi no menmen wa futate ni wakarete omotemon yoriwa Ōishi Kuranosuke o sōdaishō to nashi Horibe Yahē ika chūgi no menmen emono emono o uchifurui omotemon genkan yori rannysureba nemimi ni mizu no Kira no tsukibitora wa totsuzen kanae no waku ga gotoku kore ni ōtaisuru to iedomo chūgi ni koritaru yaiba ni wa hamukau sube naku nogaruru mono wa owazu mukau mono wa kirisutete shitsunai ni uchiittari"--Right side of image.

図2 忠臣義士四十七士玄関討入:玄関. MARC 500 field.

当大学図書館の資料のデジタル化がどんどん進み、OCRによる全文検索機能が加わった現在、通常の書籍でデジタル化されているものは当大学図書館のオンラインカタログ([SearchWorks](#))で全文検索をすることが普通になった。¹⁾しかし、テクニカルサービス部門長としての業務は目録作成業務のみではなく、新規資料の発受注、雑誌などの継続資料の受け入れ、所蔵登録、予算管理等の業務にも同様に目を配る必要があるため、一資料に対して費やせる時間はどうしても限られてしまう。確かに、利用者が資料にアクセス出来るようにメタデータを作成することは目録作成上の最重要目的であるが、一見して読めない詞書に時間をかけることは事実上困難である。「[Travel through time: Japan](#)」コレクションは17世紀から20世紀初頭にかけての日本旅行に関連する画像資料からなるコレクションであり、旅行者が集めたものや観光地で土産用に販売されていた絵地図、境内図、仏版画など、多岐にわたる資料を含む。このコレクションは当大学のDavid Rumsey Map Centerや当大学図書館のコンサベーション部門との共同展示でも数多く使用されており、注目されているコレクションである。しかし、コレクションに含まれている「[信州善光寺みやげ](#)」の詞書のように、一部のみしか解読できず有用な情報を注記フィールドに入れられないことも多くあった。(図3)



図3 信州善光寺みやげ

さらに、東亜図書館協会(CEAL)の年次総会での筆者の発表にあるように²⁾、「曲尾日記コレクション」などのようにメタデータが大量に存在する場合、現時点では当大学のオンラインカタログ上でデジタル画像に日本語を併記し、全文検索機能を付けることがまだ実現出来ないため、ソフト面だけではなくハード面の課題も重なっていた³⁾。このコレクションは、20世紀半ばに米国カリフォルニア州のオークランドとサンフランシスコで輸入事業を営んだ曲尾久雄氏(1889-1960)が書いた日記40冊と、妻の曲尾房江氏(1889-1960)が書いた日記1冊から成り、実に1920年から1960年までの期間にわたる膨大な一次資料である⁴⁾。幸いにして、量こそ多いが、現在の日本語が読める者なら容易に読める書体なので、書き起こしにはクラウドソーシングという方法も十分に考えられた。当大学図書館では2018年の時点で既にクラウドソーシングプラットフォームのFrom the Page⁵⁾やTranskribus⁶⁾などを利用して手稿資料を解読しており、成功を収めているのだが、残念ながらこれらのツールには未だ日本語処理機能が備わっていないため、「曲尾日記コレクション」には応用出来ないでいるのが現状である。

メタデータ作成の「時間」が足りない例が「曲尾日記コレクション」だとすると、さしずめ「能力」が足りないため処理が出来ない例が「[熊本藩文人書状集](#)」(以下「書状集」ということになる。そのような葛藤を抱えていた2019年の9月、カリフォルニア大学バークレー校図書館のマルラ俊江氏、白石直美氏からの紹介でバークレー校にて開催された立命館大学アート・リサーチセンターの赤間亮教授による「AIくずし字解読支援・指導システム」を使用したワークショップに思いがけなく参加することが出来た。そのワークショップ内で初めて「AIくずし字解読支援・指導システム」について知り、またワークショップ中で実際に古典籍を解読していくにあたって翻刻データベースの分かりやすいインターフェイス、カーソルで解読領域を選択するだけで瞬時にAIによる文字候補フィードバックがあらわれることに驚愕し、「読める…読めるぞ！」と感動に身を震わせていた

某大佐の姿が頭をよぎった。ワークショップが終わる頃には一つのアイデアが頭に浮かんでいた。必要最小限のメタデータしか存在しない「書状集」にこの素晴らしい翻刻データベースを使えないだろうか、と。そこからは驚くほど簡単に企画を進めることが出来た。東アジア図書館館長からプロジェクト申請の許可を得、立命館大学アート・リサーチセンターから申請が受領された直後には、当館でスキャンした画像を翻刻学習データベースに登録して頂き、データベースの使用方法を再度確認したのみで、実際のプロジェクトが開始するまで何の問題もなくあっという間にプロジェクトを開始する用意が全て整ったのだった。

第1 phase

2020年1月9日に立命館大学アート・リサーチセンター「日本文化資源デジタル・アーカイブ研究拠点 (ARC-iJAC)」2019年度 国際共同研究 申請書〔研究設備・資源活用型〕に応募した。⁷⁾スタンフォード大学東アジア図書館が所蔵する「書状集」は31枚の裏紙に貼られた書状・詩などからなるコレクションであり、購入時には書店から提供された最小限のメタデータが存在するのみであった。将来コレクションの画像が図書館のオンラインカタログに取り込まれることを見越して、画像は既にスキャンして保存済みであった。立命館大学 ARC の「AI くずし字解読支援・指導システム」を使用し、これらのデジタル画像から翻刻文を作成し、その翻刻文に基づきスタンフォード大学図書館オンラインカタログ上に公開されるメタデータの充実を目指す事を最終目標とした。当図書館の日本語部門が海外の大学との共同研究に応募するのはこれが初めての試みであった。コレクションの規模が小さいことを鑑み、このコレクションを今後のパイロットプログラムと位置付ける事にした。本稿執筆時の2022年現在となつては素人考えの浅はかさが恥ずかしい限りだが、2019年度申請当初はコレクションの規模が小さいため、短時間で全書状を解読できるのではと考えていた。ここで、「書状集」を簡単に紹介する⁸⁾。全部で109点ある書状は、全て木製の箱に納められており、購入時に書店から送られてきた情報は以下の通りである。

- 台紙に数枚を貼付した百枚を超える書状のコレクション
- 書状だけでなく、書や絵入の詩歌も含まれる
- 書状の宛名には、細川家の郡代を務めた宇野家の当主、宇野甚五郎やその嫡子宇野騏八郎の名が見られ、その他熊本藩家老有吉市郎兵衛ら宛のものもあるため時代は近世後期のものが多いと思われる

しかし、実際にプロジェクトを開始してみると解読作業は思うように進まなかった。一番の理由として挙げられるのは、プロジェクトに一定の時間を費やすことが出

来なかったことであつた。2020年4月に成果報告を提出した時にはプロジェクトが開始してから既に3ヶ月が経っていたが、共同研究申請をした当初の目的の全書状の翻刻を完了することは出来ずコレクション全体像を掴むだけで期日を迎えた。この時点で明らかになったことは下記である。

- 109点の資料のうち、漢詩が17点あつた。
- 83点の書状のうち、実に27点が宇野騏八郎氏宛の書状だった。
- 購入時に付随していた送り主の情報が間違っていることがあると判明したので(例:eal0014_07_01)、送り主の情報を全て詳しく見直す必要が出てきた。
- 「AI くずし字解読支援・指導システム」は画像、データの読み込みが共に早く、大変使いやすかった。
- 手書きのくずし字は個々人のくずし方により楷書に近い手蹟のほうが解析度が高いことが分かった。

数多くの課題が残つたため、次年度の国際共同研究に引き続き応募し、そこでは全文書の翻刻を完了することを最終目標として掲げた。また、書状内の諸事項について不明な点を問い合わせるため、熊本市の郷土史家に連絡を取るという目標を新たに加えた。

新型コロナウイルス流行によるリモートワーク下の第1 phase

新型コロナウイルス流行のため、当大学図書館は2020年3月10日をもち全館が閉館することとなった。⁹⁾ リモートワークのため、以前よりも纏まった時間を第1 phase に費やすことが出来たので、まずはくずし字の基本を復習することから開始した。「御」「候」「座」「被」「相」「申」などの頻出文字についてくずし字辞典などの例を参考にし、字形を覚えることに専念した。また、頻出部首のくずし方の例を学習し特徴を覚えることにも努めた。特にプロジェクト開始時には全く区別がつかなかった「しめすへん」「金へん」「さんずい」などのくずし方のバリエーションを、数多くの例を見ることによって理解を深めることにした。しかし、変体仮名、平仮名に関しては翻刻データベースの Deep learning 機能と字形画像検索機能を使用しても満足な解読結果が得られない状態が続いていた。これは領域選択のため文字と文字の切れ目を判断することが困難だったことによる。翻刻データベースのカーソルで領域選択をする際に、変体仮名平仮名の箇所は一文字で領域選択すべきなのか、二文字まとめて選択すべきなのか判断出来ないケースが多くあつた。それに加えて、同じ書状の中でも同一の文字が異なるくずし方をされていることも判明した。「書状集」には楷書体に近い書

体で書かれた漢詩や短い詩などもいくつか存在するため、それらの資料は翻刻データベースでの学習成果で全文の解読が容易に出来る資料であった。その一方、短い詩でも内容の意味から不明文字を推察出来ないものもあった。俳句の例を二句挙げる。作品番号「eal0014_18_1 山内對鷺の俳句と絵」では、右下には手桶に生けられた花の絵があり左側には俳句が記されている。(図 4)。翻刻データベースの支援機能を使い、「洋柿や 門方指さす 蓮の花」と解読したが、添えられた絵と、俳句との間に何らかの関係があるのかさえ、この時点では不明であった。作品番号「eal0014_24_3 野田一溪俳句」でも「雛まつり 君事通ふるは 男の子」と解読したが、俳句としては字余りであり「雛」の後に続く三文字の平仮名に対しては字形画像検索機能を使用しても対応する文字を確定することはできなかった。第 1 phase は他の書状の解読も、これらの俳句の解読状況と同レベルで停滞している状態であった。

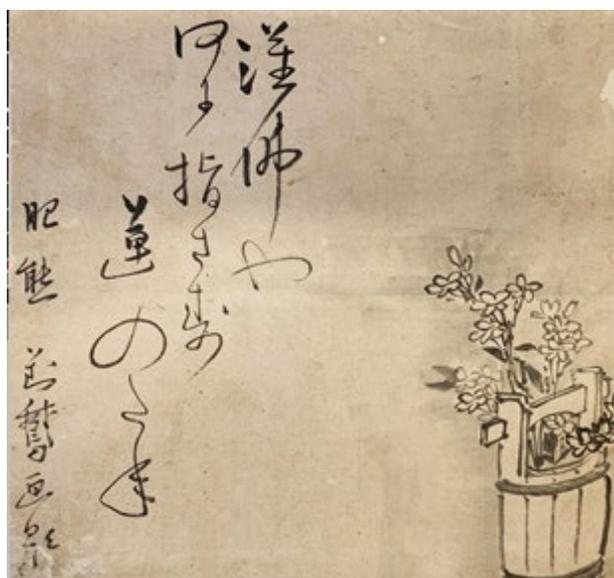


図 4 作品番号 eal0014_18_1 山内對鷺の俳句と絵

第2 phase

当大学では 2020 年 6 月からオフィス勤務が許可され、「書状集」プロジェクトに費やせる時間が再び減った。2021 年 10 月に「くずし字翻刻システムを使った古文書解読錬成講座 第 2 Phase キックオフセミナー」という通知を受け、筆者も第 2 phase に参加することとなった。第 2 phase 開始後間もなくチューターの方に 64 点を添削、翻刻して頂いた。上述の作品番号「eal0014_18_1 山内對鷺の俳句と絵」の翻刻は「灌仏や 何に指さす 蓮のたね」ということであり、挿絵は灌仏会(花まつり)の供花のようだ、ということが判明した。作品番号「eal0014_24_3 「野田一溪俳句」は「雛ひとつ 取て逃けり 男の子」と添削、翻刻して頂いた。筆者の解読では本来 5・7・5 の 17 文字のところ、平仮名の領域選択

が正しく出来ていなかったため 18 文字になっているのに加え、初句も「雛」に続く平仮名の三文字が全て間違っていた。二句は一文字も合っていなかった。第 2 phase で明らかになったのは、くずし字翻刻システムがいかにより学習者の遠隔教育に適しているかということであった。解読支援画面の下部に位置する注釈、連絡欄にはチューターへの質問を記録することが出来るのでこれを利用し回答を得た。また、添削・翻刻指導を受けた後には同画面の翻刻更新履歴を見ることにより自らの習熟度をはかることが出来た。これによって書状の基本構成について把握できるようになり書き出しや結びの決まり文句についても学んだ。

翻刻データベースの解読支援の威力が十分に発揮された例も多くある。「eal0014_30_2 堀大簡の書状」や「eal0014_01_1 松下久兵衛より武藤勝平宛書状」(図 5)ではくずし字システムの添削履歴機能表示から明らかなように、結びの文章の 1 文字目が 4 月の手紙という筆者の思い込みにより、「梅」としたことを除けば、翻刻データベースの解析がほぼ正しかった好例である。筆者による翻刻は「昼之侍要即候事 御対江戸より出分候 子留守ニ而盆肖ふ申 御残多御事御座候 然共 御面談申度事有候間 明夕明後夕之内心 御苦勞私是御出 可分御申事 梅乃朝も語以上 四月十二日 武藤勝平様 松下久兵衛」であり、チューター翻刻は「愈御清安珍重之御事 御座候頃日は御出被下 候へ共留守にて懸御目不申 御残多御事御座候 然は 御面談申度事有之候間 明夕明後夕之内午 御苦勞私宅へ御出 可被下候御頼申候 書余 緩々期貴晤候 以上 四月十二日 武藤勝平様 松下久兵衛」であった。

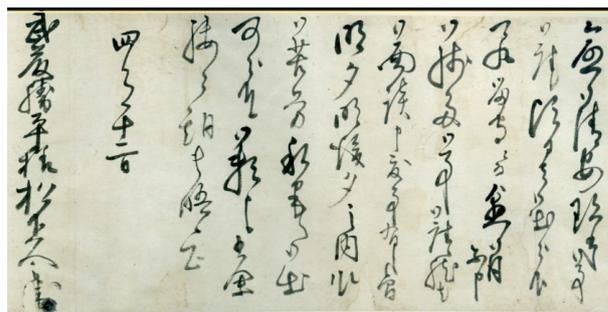


図 5 eal0014_01_1 松下久兵衛より武藤勝平宛書状

また、全ての文章ではないが、翻刻データベースの解読支援機能により部分的に高い正答率が出た例として「eal0014_05_1 小笠原唯之助より宇野甚五郎宛書状」、「eal0014_07_1 松野外記より有吉市左衛門宛書状」、「eal0014_28_2 長岡太一郎より宇野騏八郎宛書状」、「eal0014_28_4 有市宛書状」などが挙げられる。

おわりに

新型コロナウイルスの世界的大流行による思わぬ副

産物は、オンラインミーティングの機会が増え、海外との交流の敷居が驚くほど低くなったことではないだろうか。世界がつながり、情報をより気軽に交換することが可能になった。このような状況下でなければ、アート・リサーチセンターの先生方、チューターの先生方にこのように気軽に連絡を取るなど気後れしてなかなか出来ることではなかったと思う。今後の展望としては、チューターの諸先生方の添削を受け、解読を完成させた全書状に人名典拠をはじめ RDA 規則に準じたメタデータを作成し、データアクセスを強化することである。当大学のオンラインカタログで書状集の IIIF に準拠した画像データに併記して日本語のメタデータ情報を追加できるような環境が整ったときには即時使用可能なメタデータを用意しておくことを目標としたい。新型コロナウイルスが終息し、再び安全に渡航出来るようになった暁には、ぜひこれらの書状の書き手・受け取り手のゆかりの地、熊本県を巡る旅をしたいと願っている。

[注]

- 1) スタンフォード大学オンラインカタログ(SearchWorks)での全文検索については Catherine Aster 氏のブログに詳しい。Aster, Catherine A. Improved user experience for full-text search in Spotlight exhibits and image viewer. <https://library.stanford.edu/blogs/digital-library-blog/2019/07/improved-user-experience-full-text-search-spotlight-exhibits-and>, (accessed 2021-12-18).
- 2) Mazza, Mieko. Beyond Library's Catalog: Providing Metadata for Magario Family Diary Collection. <http://www.eastasianlib.org/newsite/wp-content/uploads/2018/09/20180322CEAL-CTP-03-Mazza.pdf> (accessed 2021-12-18).
- 3) Mazza, Mieko. The North American Coordinating Council on Japanese Library Resources Presentation 1: Best Practices for Metadata Creation for Digital Scholarship Projects: Through Illustrative Cases from Stanford University Libraries. <https://www.youtube.com/watch?v=A6VJZku16oc> (accessed 2021-12-18).
- 4) Magario family diaries, 1920-1972, three linear feet. https://searchworks.stanford.edu/catalog?utf8=%E2%9C%93&search_field=search&q=magario (accessed 2021-12-18).
- 5) Crowdsourcing metadata and IIIF. Special Collections Unbound. Stanford University Libraries. <https://library.stanford.edu/blogs/special-collections-unbound/2018/02/crowdsourcing-metadata-and-iiif> (accessed 2021-12-18).
- 6) Transkribus.

- <https://readcoop.eu/transkribus/>
- 7) 立命館大学アート・リサーチセンター 文部科学省国際共同利用・共同研究拠点「日本文化資源デジタル・アーカイブ研究拠点」2019 年度 国際共同研究成果報告書 [研究設備・資源活用型] E15_2019_報告書_マッツァ美恵子_web https://www.arc.ritsumei.ac.jp/lib/app/newarc/download/ijac_2019houkokusho_D16_jp.pdf (accessed 2021-12-18).
 - 8) 熊本藩文人書状集. <https://searchworks.stanford.edu/view/11585540>
 - 9) 当館の全職員がリモートワークに移行した。当館の司書は既に2015年から週に一度リモートワークをしていたが、テクニカルサービス部門は業務内容が実際の資料を手にとって作業する必要があるため、リモートワークになるということは全く考慮されておらず、よってリモートワーク対応マニュアルなども一切存在しなかった。しかし、当大学館長のリーダーシップ、University IT のサポートのもと、驚くほどスムーズにテクニカルサービス部門職員のリモートワークの準備が整い、テクニカルサービス部門の職員は自宅にデスクトップ、モニターを持ち帰って良い事になり、各自が自宅オフィスを設置、VPN で統合図書館システム(ILS), OCLC への接続環境を整え、作業開始出来るようになった。作業環境は迅速に整ったが、作業内容ではやはり当初混乱が生じた。それぞれの言語部門カタログ司書が各言語部門の職員がリモートワークで出来る作業をリストアップし職員に配分したのだが、今まで実物を手にとって作業していたため、領収書をスキャンしたものから作業したり、既存データのクリーンアップなどが主な作業として選ばれた。筆者にしてみれば、この機会に「書状集」プロジェクトにまとまった時間がとれると思ったのだが、実際には顔を合わせて職員間の業務連絡が出来なくなった分、Zoom や Slack channel を駆使して職員同士が密に連絡を取るなど、意外なことに時間が取れるということもわかってきた。

[参考文献]

- 『熊本藩年表稿』. 細川藩政史研究会, 1974.
『三百藩家臣人名辞典7』. 新人物往来社, 1989, pp. 424-469.

[謝辞]

本プロジェクトを進めるにあたり、立命館大学アート・リサーチセンターの赤間亮教授、「くずし字翻刻システムを使った古文獻解読錬成講座 第2Phase」チューターの伊東宗裕氏からご高配を賜った。プロジェクト準備段階から継続して技術サポートをして頂いた立命館大学アート・リサーチセンターの金子貴昭教授、スタンフォード大学東アジア図書館貴重書部門職員の五十嵐由美氏、立命館大学アート・リサーチセンターの職員の皆様、また、プロジェクト進行を暖かく見守ってくれたスタンフォード大学東アジア図書館のJidong Yang 館長、及び日本語部門職員の Regan Murphy Kao 博士、有田美千代氏、チャン・イーへ氏に心から感謝の意を表す。