

科学研究費助成事業 研究成果報告書

令和 4 年 6 月 30 日現在

機関番号：34315

研究種目：挑戦的研究(萌芽)

研究期間：2017～2021

課題番号：17K18495

研究課題名(和文) 日本歴史的典籍のSNS型電子テキストアーカイブ・プラットフォーム構築研究

研究課題名(英文) Research on construction of SNS-type text archive platform for Japanese historical books

研究代表者

赤間 亮 (Akama, RYO)

立命館大学・文学部・教授

研究者番号：70212412

交付決定額(研究期間全体)：(直接経費) 3,100,000円

研究成果の概要(和文)：本研究では、当初予定した「翻刻アーカイブシステム」のレベルを大きく超える、画期的な「AIくずし字解読支援・教育システム」を開発できた。本システムを使うことで、通常は、長い時間かかる「くずし字」解読能力の習得を、短期間で実現するものである。また、口頭教育の現場でも活用できるインターフェイスを備えることに成功しており、立命館大学アート・リサーチセンターの国際共同利用共同研究拠点の一機能として配置することで、本システムを多くのプロジェクトが活用できる環境も完備できた。また、この魅力的なシステムにより、国文学研究資料館や国際日本文化研究センターのコンソーシアムの一員としても貢献できることとなった。

研究成果の学術的意義や社会的意義

日本の古典籍や古文書が大量にデジタルアーカイブされ、オンライン上で閲覧が可能となり、世界中の日本学研究者が、これらの一次資料を活用して研究できる環境が確保されつつある中、記述されているテキストが「くずし字」で書かれており、現代の文字と異なるため、解読できないという日本語特有の問題が、浮び上がってきた。本研究では、このような、原文解読へのニーズを受け、初心者・中級者らでも、積極的に解読の機会を作り、かつこれらの人材の解読能力を効率よく高めていき、大規模な資料公開に匹敵する人材育成を実現させる、具体的なシステム開発である。また、閉塞感のある日本文学研究の新たな方向性を見いだす切っ掛けとなっている。

研究成果の概要(英文)：In this research, we were able to develop an epoch-making "AI Kuzushiji Decipherment Support and Education System" that far exceeds the level of the "Japanese Cursive Text Archive System" originally planned. Using this system, users can acquire the ability to decipher kuzushiji characters in a short period of time, which would normally take a long time. The system is also equipped with an interface that can be used in an educational setting, and by placing it as a function of the International Joint Research Center of the Art Research Center, we have created an environment in which many projects can utilize this system. This attractive system has also enabled us to contribute as a member of the consortium of the National Institute of Japanese Literature and the International Research Center for Japanese Studies.

研究分野：日本文学

キーワード：くずし字 古典籍 翻刻 オンライン研究会 AI 教育システム

1. 研究開始当初の背景

従来、歴史的典籍・古文書などの解読は、熟練者が翻刻作業を行い、一般向けに翻刻本文・校訂本文を提供することで、それらの作品・資料が専門家でなくとも読めるようになり、作品が普及していた。ところが、Web上で、歴史的典籍原本の画像が大量に公開されるようになり、原本にアクセスする上での地理的な障害がなくなり、世界中の研究者等が自由に原本を解読して、研究活動に、2次的にはなく、直接利用する機会が増大した。このことは、従来の翻刻本文の提供状況が、実際に現存する歴史的な典籍・文書の量に比して、まったく以て十分ではなく、くずし字を解読できない研究者や一般人にとっては、一部の(ある一定のバイアスが掛った)歴史的記述のみが、享受できていたのだという現状を、鮮やかにあぶり出すことに繋がっている。

こうした、くずし字を解読できない研究者や一般読者にとっての「原本画像閲覧」の容易化が、くずし字解読学習に対するモチベーションを大きく上げる機会となっており、また、限定された専門家のみでは遅々として進まなかった翻刻作業のペースを格段に進展させる可能性が見えてきたのである。

こうした状況の中で、国文学研究資料館がマネジメントする、「歴史的典籍の国際共同研究ネットワーク構築計画」が、歴史的典籍の大規模なデジタル化事業を推進し、また先行していた国会図書館、早稲田大学図書館らの資料とも相まって、上述のように10年前には考えられないほどの原本画像閲覧環境が、日々拡大している。

新たな方向性として本研究が目指したものは、歴史典籍のデジタル画像を使った使いやすい翻刻アーカイブシステムを提供し、研究者等がそれを活用することで、日常的な研究活動の中で、自然とテキストが蓄積させようとするものである。

本研究を開始するに当たって、まったく同時期の2017年から、「みんなで翻刻」というプロジェクトが立ち上がり、これはむしろ歴史的地震関連資料を「読む必要性に駆られて」開発したシステムであったが、SNS型モチベーション効果を取り入れた優れたシステムとして注目を集めるようになっており、オンライン上でのプロパガンダと、スマートフォン向けアプリの開発などによるインパクトにより、本研究の活動は、完全に後塵を拝するかのように見うけられた。

一方、本研究は、急激なAIの進化にも注目していた。2012年頃から、深層学習によるAIの実用的な進化は、目を見張るものがあり、とりわけ画像認識の分野では、実用的なレベルの成果が続々と生まれてきていた。これも国文学研究資料館が歴史的典籍NWプロジェクトで進めた活動であるが、くずし字の一字切出し画像の文字セットを公開し、これを利用することにより、深層学習によるくずし字認識の精度が格段に上がっていた。本研究では、くずし字OCRの研究開発で一歩進んでいた凸版印刷株式会社に協力を依頼し、AIくずし字解読システムをAPIにより提供を受けるといった方向性を見いだしていた。

こうして、自学自習型、あるいはオンライン指導型「教育システム」としての「くずし字解読システム開発へと梶を切ったのである。

2. 研究の目的

平安期以降、明治前半期までに成立した日本の歴史的典籍について、専門性の高い研究者が電子テキスト化とクラウドデータベース上でのテキスト蓄積を積極的に行えるよう、それを促進するネットワーク上のシステム環境を整え、継続的でオープンなクラウド型電子テキスト・アーカイブシステムを構築・稼働させ、オンライン上で歴史的典籍の電子テキストアーカイブが、進展していく実践事例を提案する。

開発に当たっては、国文学研究資料館を中心として進められている歴史的典籍の電子化事業との連携を意識し、この事業の次期ステップとして期待される、電子テキスト化事業に効果的なプラットフォームとして提案するものである。

また、昨今のAI技術を積極的に取り入れ、単なるテキスト・アーカイブシステムではなく、自学自習できる、あるいはオンライン型くずし字指導ができる「教育システム」としての性格をも付与し、翻刻能力のある人材を育てるシステムとする。

本研究では、方法論の確立と同時に、具体的な電子テキスト・アーカイブシステムを構築し、研究者らに、実際にこのシステムを活用してもらいながら、クラウド型アーカイブ活動が活発化するよう、改良を繰り返していく。かつ、世界標準となる日本の歴史典籍翻刻システムを実現することを目的とする。

3. 研究の方法

- (1)立命館大学アート・リサーチセンターが提供する世界最大の古典籍閲覧システムに翻刻本文記述機能を追加し、語彙へのアノテーション付与システム、原本画像と対比可能な本文閲覧システムを開発。

- (2) 翻刻支援システムの利用による WEB ブラウザを使った翻刻作業フローの確立については、凸版印刷株式会社が開発している AI くずし字解読システムを本システムに組み込み、くずし字を十分な精度をもって読める人材に限定しない、一般参加者の能力提供が可能な作業フローの確立を目指す。
- (3) 一方で、役者評判記の本文翻刻を進めている「役者評判記研究会 98」から評判記翻刻本文の提供を受け、セキュリティ機能を追加することで、公開情報をコントロールした電子テキスト活用システムを開発する。また、古浄瑠璃、絵入狂言本など、役者評判記と関連が深く、すでにデジタルテキストを作成済みの本文をシステムに搭載し、ある分野のテキストアーカイブが本システムに搭載されることにより、その分野のデジタルテキストがどれだけ研究活動に効果的な結果を齎すかについて実証実験を行う。
- (4) 加えて、著作権の消滅している活字テキストから OCR を使った本文作成を行い本システムに搭載する。活字本の画像も並置して表示することで、電子テキストの誤読部分は即座に確認できるようにする。
- (5) 院生・学部生を集め、支援システムを利用した作業を実際に行い、実験的なデジタル本文を蓄積する。なお、モニタリングにあたっては、絵画や絵入本など、作品の本文がそれほど長くなく、ビジュアル的にも学生らの興味を繋げられるものを選定する。
- (6) 日本国内では、大学、高校の授業、地域の古文書解読グループなどの場に、本システムを導入し、教育システムとしての効果をモニタリングしながら、実用レベルをアップする。
- (7) 海外の日本学研究者のくずし字解読ニーズの高まりを受け、海外の大学、研究組織へのシステム活用実証実験を行い、実用レベルを確認の上、継続的利用を慫慂する。

4. 研究成果

(1) 構築した「くずし字解読教育システム」について

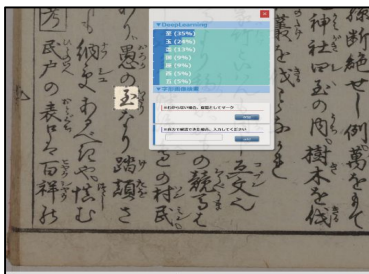
本研究の最大の成果は、「AI くずし字解読・教育システム」を完成したことである。本システムは、完全に実用的なレベルにある。以下、そのシステムの特徴について記述する。

本システムは、オンライン上の世界最大の日本古典籍データベースである、立命館大学 ARC「古典籍ポータルデータベース」の1機能として稼働している。そのため、翻刻対象にできる古典籍の数・バリエーションが最も豊富である。

「古典籍ポータルデータベース」は、立命館大学所蔵作品だけでなく、国文学研究資料館「新古典籍総合目録データベース」や、国会図書館「デジタルコレクション・古典籍資料」、早稲田大学図書館「古典籍総合データベース」などの大規模古典籍データベースは勿論、海外の博物館や図書館が所蔵する日本古典籍（主に、研究代表者が別プロジェクトでデジタル化を進めていた資料群）もまとめて検索できるため、ある一資料を翻刻しようとする場合、世界中に存在する同一書籍とページ単位で比較し、もっとも読みやすい作品を選んだり、異版を意識しながら翻刻を進めることができる。

翻刻作業は、グループや個人単位でパスワードが発行され、翻刻作業だけでなく、資料そのもののメタデータを変更・追加でき、かつ UserMemo や ImageNote という、翻刻機能以外の、便利なツールを活用することで、より効率的な協働作業を展開できる。

グループや個人の翻刻にあたっては、自分や自分のグループのみが参加できる「自分 DB」を利用することができ、オンライン上の他の研究者の活動と切り離れた翻刻作業を実施できる。もちろん、この場合でも最終的には、ポータルデータベースへの一般公開、また自分 DB の URL を公開することによるプロジェクトの成果を明確にした翻刻結果の公開ができる。



翻刻本文を書き込むことができる翻刻画面には、AI による「解読支援機能」が付いており、解読支援ボタンにより、一字単位での翻刻文字示唆を AI から受けることができる。AI 解読の能力は、完全ではないが、中級者程度の能力があるため、初心者にとっては、十分な示唆、また中級者にも効果的なアドバイスが受けられる。なお、上級者にとっても、参考となる示唆を与えられる可能性もある。

本翻刻システムは、AI の支援機能と同時に、グループ型翻刻指導システムとしても活用できるように工夫されており、指導者がいる教室での授業や、リーダーがいる研究会などで、初心者と上級者が混在しているような場でも、十分に協働作業ができるようになっている。

さらに、翻刻したテキストは、保存（データベースに格納）されると、即座に全文検索の対象となり、結果を文脈付で表示できる。これにより、同様な語彙が使われている場所などを他の作品のテキストから検索しながら、参照・参考にできる。

なお、本システムには、頁単位での付箋型メモ（UserMemo、ImageNote）機能があり、翻刻本文とは別に、索引型 DB を構築できるほか、諸本の順位付け（Version の整理）をして区別することで、別本であるが同一作品に対して何度も翻刻を施すというような無駄を省くことができる。



(2) 研究活動の経緯

【2017 年度】既存の古典籍データベースの付属していた翻刻画面を本研究のために大幅バージョンアップし、パソコン画面上で、縦書き編集ができ、画像上でのマウスを使わずにウィンドウ位置を自由に変更できる機能、ならびに文脈付き語彙検索システムを加えることにより、効率のよい翻刻作業が可能となった。また、海外の大学への、本プロジェクトの説明と、素材アーカイブを実施した。特に、米国カリフォルニア大学バークレー校とは、東アジア図書館において翻刻ワークショップを実施し、継続的協力を得られることを確認した。

【2018 年度】「翻刻本文アーカイブシステム」は、担当者・注釈・備考欄の改良を行い、さらに編集履歴を残すことが可能となった。また、翻刻本文は、オンデマンドで全文検索可能とし、KWIC 表示型が可能となった。加えて、AI 翻刻支援システムについては、検索エンジンの API 提供を凸版印刷から受け、加えて、はこだて未来大学の寺沢憲吾准教授の画像マッチングシステムを取入れた、「AI 支援付き翻刻本文アーカイブシステム」となった。「みんなで翻刻」システムと相違し、本システムが教育支援システムとしての性格があり、グループ型教育に適していることを確認した。また、翻刻済みのテキストと既存の書籍の中の翻刻本文の PDF を統合的に検索できる本文検索システムの開発にも成功した。これらの機能を踏まえて、海外の連携大学での大学院生を対象としたワークショップの形で、システムの実証実験を行なった。実施先は、カリフォルニア大学バークレー校東アジア言語学部、シカゴ大学美術史専攻、カフォルスカリ大学日本学専攻である。

【2019 年度】歴史的典籍テキストアーカイブの現場として、海外日本研究者、②大学等の高等教育機関、地域の歴史・文学愛好者という3つのフィールドをターゲットに「AI くずし字解読教育・支援システム」を実用的に活用してバージョンアップを行った。4月から7月には、立命館大学文学部の講義において本システムを使った授業を実施。5月、蘭国・ライデン大学の国際研究会、8月、英国・ケンブリッジ大学の日本古典籍サマースクールの講義枠、9月、米国・UC バークレー校でワークショップを開催し、学内学生・研究者だけでなく、周辺大学の研究者の参加もあった。11月には、日本近世文学会において、授業での実践結果の報告を行ない、一定の評価を得た。12月・1月には、地域勉強会への提供事例として、舞鶴市において、地域型書籍コレクションを使ったワークショップを実施。高校生から寺社などの文化財所有者、一般市民、周辺自治体からも参加者を得た。なお、舞鶴市のイベントは、文化庁の「大学・研究機関等との共同研究事業」と共同し、地域文化の活動の創生を目的に掲げたものである。

「解読支援」システムは、ARC データベース群の内、浮世絵、古典籍、文化資源、番付への登載を完了。さまざま歴史的典籍のすべてに対応できるようになった。また、翻刻対象典籍が IIIF で公開されている場合ドラッグ&ドロップで、簡単に ARC データベースへの取り込みが可能となり、対象となる古典籍・文化資源は無限となった。ARC ヴァーチャル Ins. に「古典籍デジタル研究所」を開設し、「古典籍翻刻プロジェクト」の中で DEMO プロジェクトを設置し、複数の独立する翻刻プロジェクトへの対応が可能となった。

【2020 年度】「AI くずし字解読教育・支援システム」として本格的な普及・実証実験が可能な段階にあること、研究代表者が学外研修期間であったことを踏まえ、欧米各国でのワークショップの複数回実施を予定していたが、コロナ禍の影響により一切開催ができなかった。一方、国内では、舞鶴工業高等専門学校での授業実践が前期後期において実施され、理系生徒・学生に対する国語・文字教育における効果が実証できた。この授業を受けて、舞鶴市の文化財を使った市民向けのくずし字講座を実施したが、授業で使用方法を身につけた学生が、講座の補助にまわるなどの副次的な効果もあった。また、ポストコロナ時代を見据えたインターフェイスや細かな使用感の向上を狙ったシステムのバージョンアップを実施した。より柔軟に授業・講座の準備ができるよう教材選択機能を追加するなどの改良を行い、また利用マニュアルのオンライン公開・翻刻規則の公開など、指導者側の環境を整え、オンラインでのくずし字講座・授業がスムーズに実施できるようになった。加えて、初級者向けの教育素材の充実を図り、システムに搭載した。

【2021 年度】コロナ禍は継続したが、本研究では、オンライン利用型システムを使った研究実践であり、「くずし字翻刻教育・指導システム」が、まさにオンライン型教育・研究時代に効果的に機能した。とりわけ人文学研究者にオンライン環境の活用が定着したがために、世界規模のより広い範囲での活用が可能となり、まず、国文学研究資料館が取りまとめるコンソーシアムの人的ネットワークを活用して、前期に初級・中級レベルの参加者を募集。「くずし字錬成講座」を海外向け2回、国内向け2回の合計4回実施し、合計80名もの参加者を得た。ここでは、主にシステムの利用方法を説明し、オンラインで添削指導が行えるチューターを雇用することで、AI と熟練者の双方のサポートによって進められるくずし字習得・翻刻アーカイブのスタイルを実現した。一度切りで終る従来の対面型くずし字講座ではなく、オンラインシステムを利用したオンデマンド型の講座が実現できることも証明した。後期には、国際日本文化研究センターの協力を得て、本格的にチューターがサポートする「オンライン翻刻プログラム」を10月から3月まで実施し、解読能力の向上とともに、多くの翻刻テキストを蓄積することができた。

なお、今年度も効果的な教材を追加購入し初心者・中級者らの教育環境を整えた。

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 2件 / うち国際共著 1件 / うちオープンアクセス 0件）

1. 著者名 赤間 亮、岡敏生	4. 巻 31.2
2. 論文標題 AI技術を応用したくずし字翻刻学習・指導システム	5. 発行年 2020年
3. 雑誌名 月刊画像ラボ	6. 最初と最後の頁 20-25
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 マルラ俊江、赤間 亮	4. 巻 -
2. 論文標題 カリフォルニア大学バークレー校 C.V. スター東アジア図書館所蔵「家伝集」のオンライン公開と利便性を高める工夫 - 国際的活用と資料館連携を目指して	5. 発行年 2019年
3. 雑誌名 人文科学とコンピュータシンポジウム論文集2019	6. 最初と最後の頁 197-204
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 赤間 亮	4. 巻 -
2. 論文標題 知の蓄積を加速させる外部データベースの育成	5. 発行年 2019年
3. 雑誌名 2019年度アート・ドキュメンテーション学会年次大会予稿集	6. 最初と最後の頁 6-7
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計7件（うち招待講演 2件 / うち国際学会 5件）

1. 発表者名 赤間 亮	
2. 発表標題 芸術文化資源デジタル・アーカイブの国際的共同利用（趣旨説明） - オンライン環境での知的生産システムとそのツール	
3. 学会等名 アート・ドキュメンテーション学会	
4. 発表年 2020年	

1. 発表者名 赤間 亮
2. 発表標題 AIくずし字解読支援機能付翻刻システムによるくずし字指導の実践と活用提案
3. 学会等名 第137回 日本近世文学学会大会
4. 発表年 2019年

1. 発表者名 赤間亮
2. 発表標題 如何為藝術建置數位資料庫 由立命館大學藝術研究中心的經驗
3. 学会等名 超越時、空與文化之數位資料庫建置與活用 (台湾芸術大学) (招待講演) (国際学会)
4. 発表年 2018年

1. 発表者名 Ryo Akama, Ryoko Matsuba
2. 発表標題 The Art Research Center 's Digital Archives and Research Environment
3. 学会等名 Digital Japan Mini-Festival (Sainsbury Institute of study for Japanese Art and Culture) (招待講演) (国際学会)
4. 発表年 2018年

1. 発表者名 RYO Akama
2. 発表標題 The Art Research Center 's Digital Archive System
3. 学会等名 On the Digital Archive and Its Uses for Japanese Humanities (国際学会)
4. 発表年 2018年

1. 発表者名 RYO Akama
2. 発表標題 Japanese Woodblock Prints of Tales and Plays in Landscape Format
3. 学会等名 Landscapes in Art, Theory, and Practice across Media, Time, and Place (国際学会)
4. 発表年 2017年

1. 発表者名 RYO AKAMA
2. 発表標題 Promoting online resources to enhance understanding of Edo-period illustrated books Developing Virtual Exhibitions
3. 学会等名 16th International Conference of the European Association of Japanese Studies (国際学会)
4. 発表年 2021年

〔図書〕 計1件

1. 著者名 村上征勝、金明哲、小木曾智信、中園聡、矢野桂司、赤間亮、阪田真己子、宝珍輝尚、芳沢光雄、渡辺美智子、足立浩平	4. 発行年 2019年
2. 出版社 勉誠出版	5. 総ページ数 850
3. 書名 文化情報学事典	

〔産業財産権〕

〔その他〕

<p>ARC古典籍ポータルデータベース https://www.dh-jac.net/db1/books/search_portal.php 日本古典籍デジタル研究所 (古典籍翻刻プロジェクト) https://www.arc.ritsumei.ac.jp/lib/vm/J-book/A/ 浮世絵ポータルデータベース https://www.dh-jac.net/db/nishikie/search_portal.php 古典籍ポータルデータベース https://www.dh-jac.net/db1/books/search_portal.php 番付ポータルデータベース https://www.dh-jac.net/db1/ban/search_portal.php 文化資源ポータルデータベース https://www.dh-jac.net/db1/resource/search_portal.php 古典籍デジタル研究所 https://www.arc.ritsumei.ac.jp/lib/vm/J-book/ ARC A Simple Text Archive https://www.arc.ritsumei.ac.jp/text_archive/ ARC古典籍ポータルデータベース http://www.dh-jac.net/db1/books/search_portal.php ARC古典籍・全文テキスト検索 http://www.dh-jac.net/db1/books/search3.php ARC古典籍ポータルデータベース (翻刻システム) http://www.dh-jac.net/db1/books/search_portal.php</p>

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計4件

国際研究集会 Printed Books and Knowledge about Necessities and Luxuries in Early Modern China and Japan	開催年 2019年～2019年
国際研究集会 AIでくずし字を読む。－くずし字解読支援システムの紹介と、くずし字解読ワークショップ「くずし字駅伝」(Ca' Foscari University of Venice)	開催年 2018年～2018年
国際研究集会 くずし字錬成講座 Phase 1	開催年 2021年～2021年
国際研究集会 くずし字錬成講座 Phase 2	開催年 2021年～2022年

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関			
米国	カリフォルニア大学バークレー校	シカゴ美術館	シカゴ大学	他2機関
英国	ケンブリッジ大学	大英博物館	セインズベリー日本藝術研究所	
オランダ	ライデン大学	ライデン民族学博物館		
米国	シカゴ大学	カリフォルニア大学バークレー校		
イタリア	カフォスカリ大学			
ドイツ	ベルリン自由大学			
スイス	チューリッヒ大学			
ベルギー	ルーベン大学	ベルギー王立歴史美術館		