

くずし字学習支援アプリケーションの開発

橋本 雄太¹ 久田 行雄² 有澤 知世² 小林ベター ダニエル² 飯倉 洋一³

概要: 文化資料デジタルアーカイブの普及, またそのオープンデータ化の進展によって, 日本語の古典籍資料や歴史資料に対するアクセスは近年飛躍的に向上している. これらの資料を利用した国際的共同研究や, 異分野融合型研究の実現が期待されているが, 日本語資料を伝統的諸分野の枠外で活用するノウハウや仕組みは未だ発展途上である. とりわけ前近代の古典籍・歴史資料は, その大多数が現代では使用されないくずし字や変体仮名で記述されており, 専門家以外には解読すらままならないという現状がある. 発表者は, くずし字解読の学習支援モバイルアプリケーションである「KuLA」を開発し, 2016年2月から公開している. KuLAは(1)字形学習モジュール, (2)読解演習モジュール, (3)コミュニティモジュールから構成され, 近世の刊本から採取した3,000件以上の文字画像と3点の翻刻文付き読解教材を教材として収録している. これらの教材を利用してくずし字解読技術の習得を支援するだけでなく, 難読文字を他ユーザーとのコミュニケーションから集合的に解決する機構をKuLAは併せ持っている. 公開から2ヶ月が経過した4月現在までにKuLAは13,000人以上にダウンロードされており, 国外の日本古語教育の現場でも実用に使われている.

キーワード: くずし字, 変体仮名, 古典籍, 歴史資料, モバイルアプリケーション, 集合知

Developing a Learning Application for Reading *Kuzushiji*

YUTA HASHIMOTO¹ YUKIO HISADA² TOMOYO ARISAWA² DANIEL KOBAYASHI-BETTER²
YOICHI IKURA³

Abstract: The past several years have seen the rapid improvement of access to digitized Japanese classical books and historical manuscripts. It is highly expected that the international and interdisciplinary research collaborations occur based on these materials. However, academic uses of these materials outside traditional disciplines are still in development. In particular, studying pre-modern Japanese manuscripts and books require the skills to read *kuzushiji* and *hentaigana*. The aim of KuLA, a mobile application we have developed and published since February 2015, is to provide accessible resources for learning how to read *kuzushiji* and *hentaigana*. It consists of three different modules: (1) character learning module, (2) reading module, and (3) community module.

Keywords: classical Japanese literature, historical manuscripts, mobile application, collective knowledge

1. はじめに

文化資料デジタルアーカイブの普及, またそのオープンデータ化の進展によって, 日本語の古典籍資料や歴史資料に対するアクセスは近年飛躍的に向上している. 国立国会図書館デジタルコレクションでは, 数万点におよぶ古典籍

のデジタル画像がオープンデータ公開されている. 国文学研究資料館(国文研)では, 同館および連携機関が所蔵する30万点の古典籍をデジタル公開する事業が現在進行している. これらの資料を利用した国際的共同研究や, 異分野融合型研究の実現が期待されているが, 日本語資料を伝統的諸分野の枠外で活用するノウハウや仕組みは, 未だ発展途上である.

これまで, くずし字で記述された古典籍や歴史資料の原資料はもっぱら, 資料に直接的アクセスを有する日本国内の文学研究者や歴史研究者の手に任されてきた. しかしな

¹ 京都大学大学院文学研究科 博士後期課程
Graduate School of Letters, Kyoto University

² 大阪大学大学院文学研究科 博士後期課程
Graduate School of Letters, Osaka University

³ 大阪大学大学院文学研究科
Graduate School of Letters, Osaka University

から近年、日本国外の研究者や、自然科学を含む異分野からもこれらの資料が研究対象として注目されている。たとえば英ケンブリッジ大学や独ハイデルベルク大学の日本研究学科では、近年くずし字の解読講座が設けられるようになった。また、総研大で実施されている『オーロラ 4D』プロジェクトは、上賀茂神社に収蔵されている古記録等から過去の天体現象の解明を目指している^{*1}。また [1] は、過去の海洋生物の生態解明に江戸時代の本草学資料を活用したことで話題になった。

古典籍や歴史資料を活用した国際的・分野横断的研究は今後も増加するものと思われるが、最大の障壁は古典籍の活用にくずし字の解読能力が必要になることである。とりわけ近世以前の日本語の古典籍・歴史資料は、現代では使用されにくずし字や変体仮名で記述されており、訓練を受けない限り現在の日本人はほとんど解読することができない。国文学者の中野三敏は、人文系の学部であっても、くずし字の判読能力を有する教員は2割以下であろうと述べている [2]。もちろん源氏物語などの重要作品は翻刻が出版されているが、たとえば国書総目録に記載されている古典籍のうち、翻刻のあるものは一割に満たないと言われている^{*2}。したがって近世以前の古典籍・歴史資料の研究にはくずし字の解読能力が不可欠である。

情報技術を使用して、くずし字の解読を支援しようという試みも一部で進められている。凸版印刷株式会社は、くずし字の OCR エンジンの開発に取り組み、80%の識別性能を得たと2015年7月に発表した^{*3}。また、2016年3月には奈良文化財研究所と東京大学史料編纂所が共同で開発したくずし字解読システム『MOJIZO』が発表された^{*4}。これらは古典籍・歴史資料の研究に際して非常に有用なツールになり得るが、使用にあたってはくずし字解読の能力がやはり必要である。

くずし字解読の能力それ自体を習得する手段としては、国文学や史学分野の研究者によって多くの教科書が著されてきた。他方で、紙媒体の出版物には様々な制約が存在する。出版社が介在する以上、教材は有償にならざるを得ず、また日本国外からの利用にも多大なコストがかかる。教材利用の敷居を下げるためには、インターネット上で教材を配布することが最適である。特に、教材をモバイルアプリケーションとして開発すれば、利用にかかる物理的制約をほとんど解消することが可能になると同時に、紙媒体の教材では実現し得ない応答性も実現することができる。

UCLA と早稲田大学が共同で開発した『変体仮名あぶ



図 1: 文字詳細画面

図 2: テスト画面

り』^{*5} は、モバイルアプリケーションとして開発された恐らく初めての変体仮名の学習教材である。ただし現時点で『変体仮名あぶり』は変体仮名の単字の学習のみに対象が限定されており、現時点で連綿した文字や漢字はサポートしていない。

くずし字解読の能力は、字形知識の習得だけでなく、実際の古典籍や歴史資料の解読訓練も含めた総合的な学習が必要である。筆者らは、(1) くずし字の字形学習、(2) 実際の古典籍を用いた読解訓練、(3) ネットワークを利用した利用者間の相互学習、以上3つの機能をサポートする、くずし字学習の総合支援アプリケーション「KuLA」を開発し、2016年2月から公開している。本稿では、KuLA がサポートする機能と、KuLA の公開後利用者から得られた反応について述べる。

2. 実装した諸機能

2.1 字形学習モジュール

このモジュールは、くずし字解読の基礎となる字形知識の習得を目的としている。変体仮名??文字と草書体漢字??文字について、単字と連綿した複数文字の用例画像およそ3,000点を収録している。各文字は20文字前後のセクションに分割されており、細かい単位で学習を実施することが可能である。文字ごとの詳細表示画面では、図1のように用例画像とその解説が表示される。また、学習の進捗を確認するため、図2セクション単位で用例画像を元にテストを実施する機能が備えられている。

2.2 読解訓練モジュール

このモジュールでは、字形学習モジュールで習得した知識を活かし、実際の古典籍を用いた読解訓練をおこなう。現在のバージョンでは教材として『方丈記』(江戸時代に刊

*1 <https://aurora4d.jp/>

*2 この数字はあくまで推測によるもので、厳密な統計は存在しない。

*3 http://www.toppan.co.jp/news/2015/07/newsrelease150703_2.html。ただし、80%という精度は比較的判読が簡単な刊本の解読の場合で計測された数字と思われる。

*4 <http://mojizo.nabunken.go.jp/>。

*5 <http://alcvps.cdh.ucla.edu/support/>。



図 3: 読解訓練機能 (方丈記)

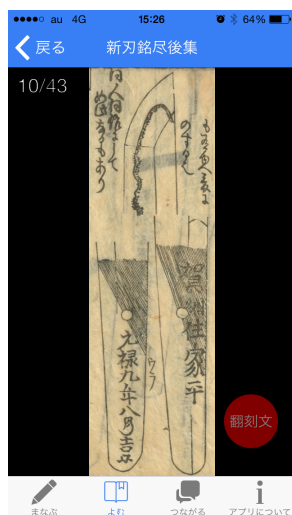


図 4: 読解訓練機能 (新刃銘尽後集)

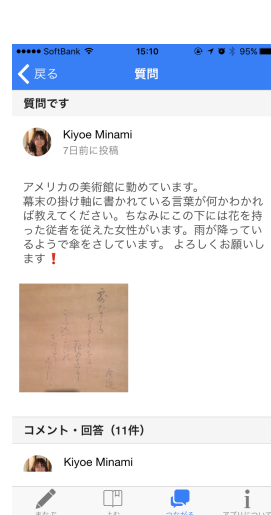


図 5: コミュニティモジュールに投稿された質問

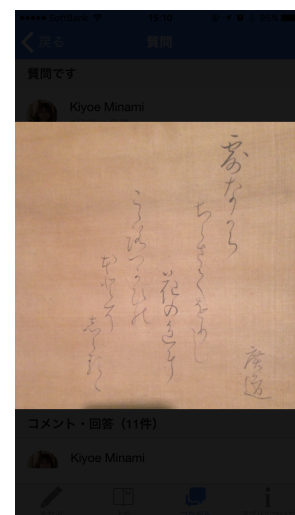


図 6: 質問に添付された画像

行された古活字版),『なぞなぞ絵双六』,『新刃銘尽後集』の3点を収録している。スマートフォンの小さい画面でも資料が容易に読めるように、一度に表示されるテキストは2~3行に制限されている(図3,4)。また、初期状態では表示されないが、翻刻文を確認したい時には資料画像の横に翻刻を表示させることができる。

2.3 コミュニティモジュール

実際の古典籍や歴史資料の読解に取り組む過程で、難読文字に遭遇することは避けられない。大学の演習等対面式の教育を受けている場合は、熟練者に質問をすることができるが、独学でくずし字を学習している場合は熟練者が周囲に存在しない場合が多い。コミュニティモジュールは、独習者でもネットワークを通じて他の学習者や熟練者とコミュニケーションを取ることを可能にし、集合知によってくずし字解読を支援することを目的にしている。

コミュニティモジュールでは、スマートフォンやタブレット付属のカメラを用いて難読文字の画像(4枚まで)を撮影し、その読み方を他ユーザーに質問することができる(図5,6)。投稿された質問は全ユーザーに向けて開示され、他ユーザーはその投稿にコメントをつけることで質問に回答することができる。

3. 公開後のユーザー反応

本研究の主眼はくずし字学習の総合的な支援ツール開発であり、既存の研究との定量的な比較やそれに基づく評価が難しい。本節では、くずし字学習者に対するKuLAの寄与について評価するため、App Store上のユーザーレビューなど、アプリケーション公開後のユーザーからの反応について記述する。

3.1 ダウンロード数とユーザー評価

KuLAは2016年2月18日にApp StoreおよびGoogle Playで同時公開された。同年4月18日現在までの2ヶ月間のダウンロード数は(iOS版で8,750, Android版が5,359)である。ユーザーからの評価は4月18日時点でiOS版が4.5/5.0(評価数29), Android版が4.7/5.0(評価数47)となっている。

また、全体のダウンロード数の約7%が日本国外からのものであり、米国、ドイツ、韓国、イタリア、イスラエルの順にダウンロードが多い。なお、イタリア・ローマ大学サピエンツァ校の日本古文書学講座にて、自習用教材としてKuLAが活用されているという報告を頂いている^{*6}。

3.2 コミュニティモジュールへの投稿

コミュニティモジュールでは、4月18日までに24件の質問が投稿された。このうち2件は筆者らによる投稿で、純粋に一般ユーザーから寄せられた質問は22件である。各質問に寄せられたコメントの数は100件を超えている。設計時の期待通り、難読文字の読み方を質問するものである。例えば図7,8のユーザーは、保有する掛け軸に書かれた文字の読み方を画像付きで質問しており、別ユーザーからの回答コメントによって難読文字が解決している。これは設計時の狙い通り、ユーザー間のコミュニケーションを通じて難読文字の「教え合い」が発生していることを示すものである。4月18日現在までに投稿された22件の質問のうち、15件がこうした難読文字に関する質問である。

現在、コミュニティモジュールの使用にはFacebookの認証が必須である。しかし実名制のFacebookのアカウントの利用は投稿の心理的ハードルを高めているものと思わ

^{*6} この報告は「Facebook上の公開グループ”Digital Humanities in Japan”」上で現地の教員から頂いた。<https://www.facebook.com/groups/758758500904522/search/?query=kula>。

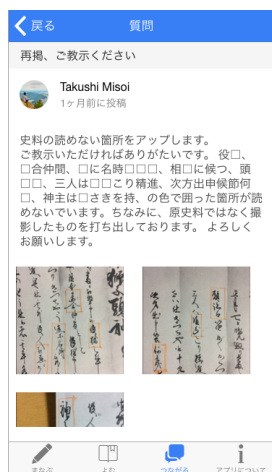


図 7: コミュニティモジュールに寄せられた質問



図 8: 同質問に寄せられた回答コメント

参考文献

- [1] Hayashi, R.: Past biodiversity: Historical Japanese illustrations document the distribution of whales and their epibiotic barnacles, *Ecological Indicators*, Vol. 45, pp. 687-691 (2014).
- [2] 三敏, 中野: 和本のすすめ: 江を読み解くために, 岩波書店 (2011).

れる．このため次回のアプリケーション更新時には Google や Twitter アカウントでの認証を導入する予定である．これが実現すれば，コミュニティモジュールへの参加ユーザーは更に増大するものと思われる．

4. おわりに

現時点では KuLA は初学者用の入門教材であり，本格的なくずし字の解読能力習得には実際の古典籍・歴史資料を用いた膨大な訓練が必要になる．しかしながら，KuLA の開発は現在も継続中であり，今後のアプリケーション更新によって国文学・日本史学の研究者の利用にも耐える教材となることを目標としている．現時点で追加が予定されている解説や機能は次の通りである．

- (1) 漢字の部首ごとのくずし方のパターンに関する解説 (字形学習モジュール)
- (2) 連綿した文字画像によるテスト (字形学習モジュール)
- (3) 連綿した文字画像によるテスト (字形学習モジュール)
- (4) 変体仮名・くずし字の総合テスト (字形学習モジュール)
- (5) PC から利用可能な Web インターフェイスの構築 (コミュニティモジュール)
- (6) 質問投稿の相互評価 (コミュニティモジュール)

先述の通り，日本語の筆記体系は明治時代以降大きな変化を遂げており，15-16 世紀の出版物の解読にも困難が生じない欧米とは状況が大きく異なる．文化資料のデジタル化が相当程度進行した現在，人文学におけるコンピューター利用の研究分野においても上記の問題が強く意識され，今後重点的な取り組みが成されるべきだと思われる．

謝辞

本研究は科研費・挑戦的萌芽研究 (15K12850) の助成を受けたものである．