

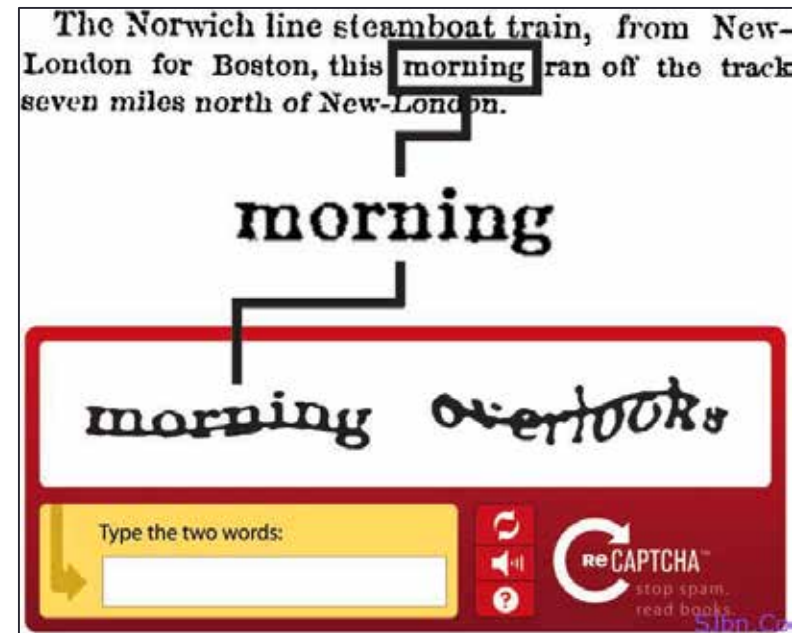
くずし字の学習支援と 市民参加翻刻

京都大学文学研究科 博士後期課程
橋本雄太

2017年2月10日@第2回CODHセミナー

はじめに

- 膨大な印刷 / 手書き資料をテキスト化する手法には、
「**多数の人々の力を活用するアプローチ**」もある
 - e.g. Google の reCAPTCHA
 - e.g. UCL の Transcribe Bentham
- そうしたアプローチの例として、
本発表では「**みんなで翻刻**」
プロジェクトを紹介





みんなで翻刻

【地震史料】

古文書・古記録を解読して、
歴史災害研究に参加しよう！

参加する

現在はPCでのアクセスのみ対応しています。

いいね！ ツイート G+ B! ブックマーク LINEで送る



京都大学
KYOTO UNIVERSITY

古地震研究会

URL: <https://honkoku.org/>



みんなで翻刻【地震史料】



はじめに

ホーム

ごあいさつ

ブログ・SNS・ML

地震史料を翻刻する

翻刻ガイドライン

史料一覧

お気に入り・作業履歴

フォーラム

くずし字解読を学ぶ

まなぶ

つながる

ライセンス/クレジット

お問い合わせ

☐ 「添削希望」のみ表示 ☐ コメント有のみ表示

お知らせ：[フォーラム](#)が設置されました。不具合や要望をお寄せください（2017-02-04）

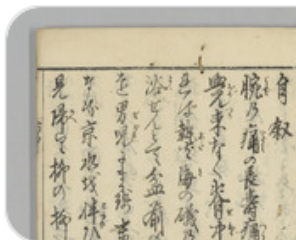
お知らせ：翻刻が修正された際に通知が届くようになりました（2017-02-05）



小林じゃさんが1文字翻刻しました

24分前

添削希望



熱海温泉圖彙（見返し題）—コマ24

本堂の壁（かべ）に詩（し）を題（だい）して住職（ぢうしよく）を雲居国師に譲（ゆず）り裏門よりい...



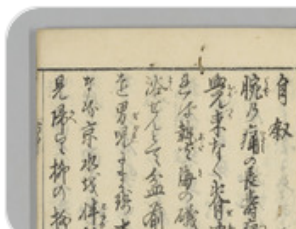
あっぱれ！



小林じゃさんが587文字翻刻しました

29分前

添削希望



熱海温泉圖彙（見返し題）—コマ26

▲紅葉狩（もみぢがり）▲鹿間（しかきく）▲蟲聞（むしきく） 暖地なればむし多し早くいでゝ遅（おそく）...

翻刻文字数ランキング

- 三上正昭
59525文字
- 小林じゃ
54330文字
- TakanoKiyoshi
52017文字
- ramayakat
40710文字
- chinjuh
40167文字
- 3dmiCg8JwW
39678文字
- 田原輝男
25476文字
- Silverback Silenos
21665文字
- なかに一
20559文字
- 倉本京子
20455文字
- antar skuita
19324文字
- Yoko Odagi
17605文字

「みんなで翻刻 地震史料」

- 概要

- 歴史災害史料の市民参加型翻刻プロジェクト
- いわゆる「クラウドソーシング」翻刻
- 2017年1月10日に Web サイトを正式公開

- 開発・運営・・・京都大学古地震研究会

- 京大理学研究科を中心として、歴史地震史料の解読にあたっている学際グループ
- 主催：中西一郎（理学研究科）、加納靖之（防災研）

プロジェクトの目的

- 古地震研究としての目的

1. 膨大な量の地震史料を短期間でテキスト化し、
地震・防災研究の基礎資料の整備に寄与する
2. 市民の歴史災害に対する意識の向上

- 人文情報学研究としての目的

- 難度の高い前近代日本語史料の翻刻に対しても、
市民参加型研究が可能であることを実証する

対象史料

- 東大地震研図書室「石本文庫」
 - 東京大学地震研2代目所長・石本巳四雄(1893-1940)が収集した災害史料コレクション
 - このうち、現在地震研でデジタル公開されている史料114点(画像数3,193枚)の全文翻刻が当面の目標



『信州浅間山焼』



『地震年代記』



『元禄地震記』 7

「学習」ベースのプロジェクト設計

- 市民参加を実現する上での課題
 - 多数の市民の誘導
 - 適切なモチベーション設計
 - 品質の担保
- 「みんなで翻刻」のアプローチ
 - 翻刻作業への参加を通じて、参加者が古文書解読の訓練を受けられるようにサイトを設計する
 - 参加者の個別的学習と協調的学習を支援
 - 「オープンエデュケーション」型クラウドソーシング(仮称)

くずし字学習支援アプリ「KuLA」との連携

概要

- 阪大文学研究科を中心に開発された「くずし字」の学習支援アプリケーション
 - 3,000枚の用例画像と読解教材を収録
 - 2016年2月の公開後 5.6万回DL
-
- 「みんなで翻刻」との連携
 - KuLA の全コンテンツをサイトに転載
 - くずし字解読の基礎を独習可能に



参加者間の相互添削システム

- 翻刻作業の完了時，
翻刻文の添削を他の参加者に依頼することができる
- 相互添削の目的
 - 参加者のスキルアップ
 - 翻刻文の品質の向上

編集を終了

この内容で保存していいかな？
すべての文字が翻刻できていたら「翻刻完了」にチェックをつけてね。読めない箇所が残っていても大丈夫！
編集コメントを記入しておく、後の人の助けになるかも！



添削おねがいします

☐ 翻刻完了

☒ タイムラインで共有

☒ 添削希望（他の参加者に翻刻文の添削をお願いします）

キャンセル 保存

翻刻作業のタイムライン表示



みんなで翻刻【地震史料】

はじめに

- ホーム
- ごあいさつ
- ブログ・SNS・ML

地震史料を翻刻する

- 翻刻ガイドライン
- 史料一覧
- お気に入り・作業履歴
- フォーラム

くずし字解読を学ぶ

- まなぶ
- つながる

ライセンス/クレジット

お問い合わせ



田原輝男さんが318文字翻刻しました
1時間前

添削希望



浅間大焼騒動記 中一コマ26

も勲れはて夫より種月然へ押寄大勢ふミ入れ共人害人も見
へず唯本堂計ぞおわしける然共委...

あっぱれ！



ramayakatさんが424文字翻刻しました
2時間前

添削希望



江戸大地震并ニ出火場所細見録一コマ3

一新吉原江戸町京町角町揚屋町伏見丁仲の丁すべてくるはの
こらずゆりつぶれ《？：てみへ出...

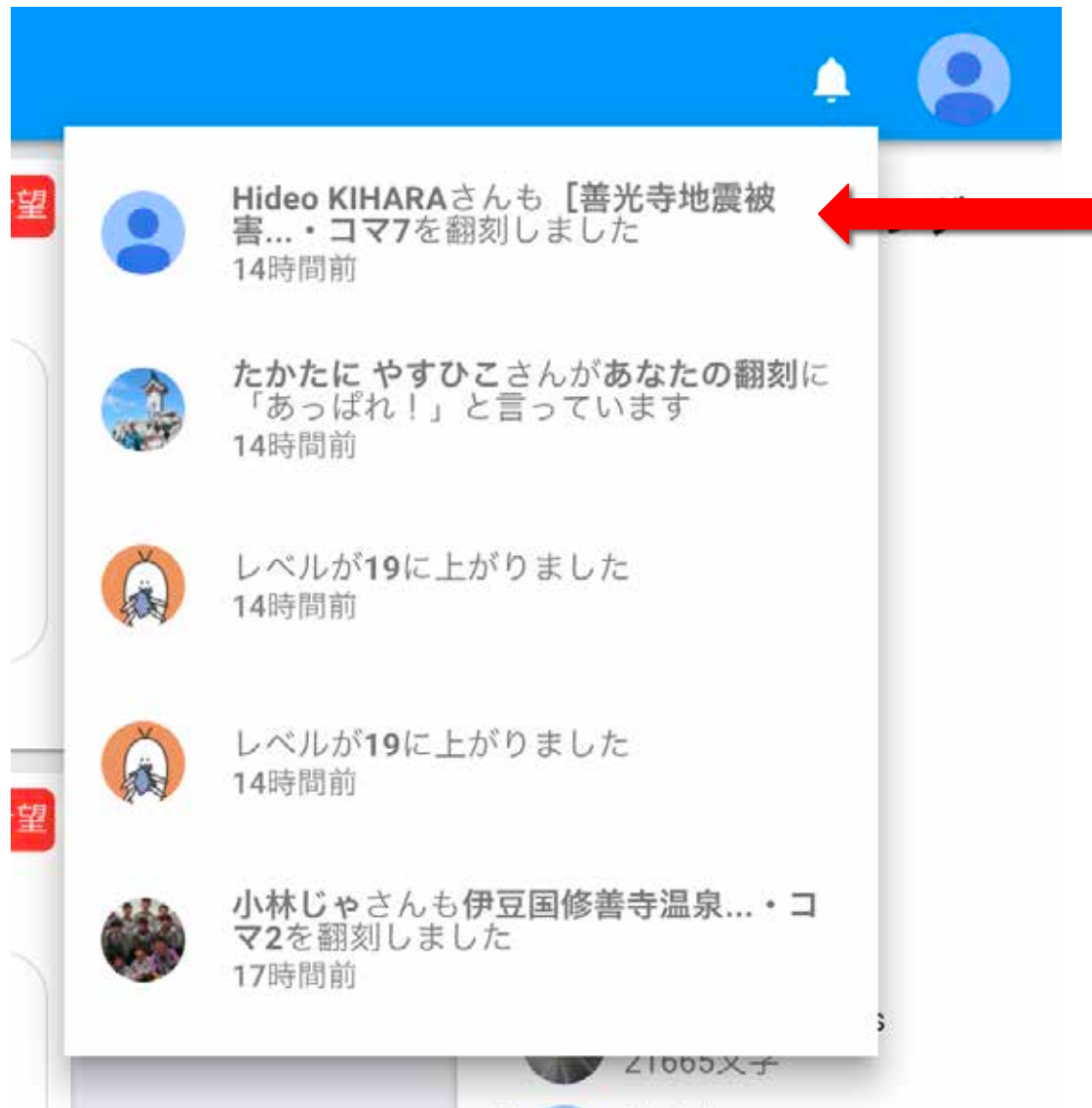
あっぱれ！



小林じゃさんが679文字翻刻しました
2時間前

添削希望

翻刻文修正時の通知



編集履歴の差分(diff)表示

地震年代記 ★ 14

閲覧 入力 ノート **編集履歴** 凡例

seizai
2017-01-13 08:20:03 - 0文字を翻刻

十一
2017-01-13 06:57:26 - 4文字を翻刻

seizai
2017-01-13 06:50:17 - 626文字を翻刻

大河
2017-01-12 11:09:18 - 130文字を翻刻

十一
2017-01-12 10:49:03 - 60文字を翻刻

大河
2017-01-12 09:48:43 - 12文字を翻刻

十一
2017-01-12 04:19:34 - 93文字を翻刻

steve forrest
2017-01-12 06:09:10 - 40文字を翻刻

十二月廿九日同前《箱…合運図》《箱…明應》二年十月同前《箱…續史
畧》同三年五月七日
同前《箱…合運図》同四年八月十五日、一鎌(かま)くら一大佛地震(だ
いぶつじしん)にて一破損(はそん)。同七年六月
十一日、一天(あめ)の一下(した)一時に大地震。以上《箱…本朝年代
記》同年八月十九日地大に震に此月
しばしば一震(ふる)ひ、一伊勢(いせ)一紀伊(きい)参一河(かは)
の一海(うみ)あふるゝ。《箱…和長郷記》《箱…天正年代記》《箱…後
柏原天皇》
《箱…永正》七年八月七日、一地震(ちしん)大にとゞまらず、一天王寺
(てんわうじ)一石(いし)の一鳥居(とりゐ)一額(がく)一落(お
ち)、一五畿内(ごきない)
一堂塔(どうたう)一破懷(やれくづ)【懷は壞の誤字でしょうか】(や
れくづ)れ一人(ひと)一多(おほ)く死す。《箱…本朝年代記》同八日
地大に震(ふる)ひ廿八日 遠江(とほ)の
一海(うみ)あふれ、陸地(くがち)一没(おちいり)て海となる。一今
切(いまぎれ)のわたしこれなり。《箱…續史畧》同九年
六月、一大地震(おほぢしん)《箱…合運図》同十一年四月八日同前
《箱…後奈良天皇》《箱…弘治》二年二月
一地(ち)大に震ふ。同三年七月同前《箱…公卿補任》《箱…正親町天
皇》《箱…天正》十三年十一月
十九日《割書…合運図》に廿九日とす《京師(みやこ)地(ち)大(おほ
い)に震(ふる)ひ百日餘(よ)やまず《箱…續史畧》《箱…後陽成天
皇》
《箱…天正》十七年二月駿河(するが)遠江(とほ)を懷【壞の誤字で
しょうか】(くづ)す。《箱…武徳大成》

差分を見ることで、どの箇所が修正されたか簡単に確認できる

公開1ヶ月後の経過

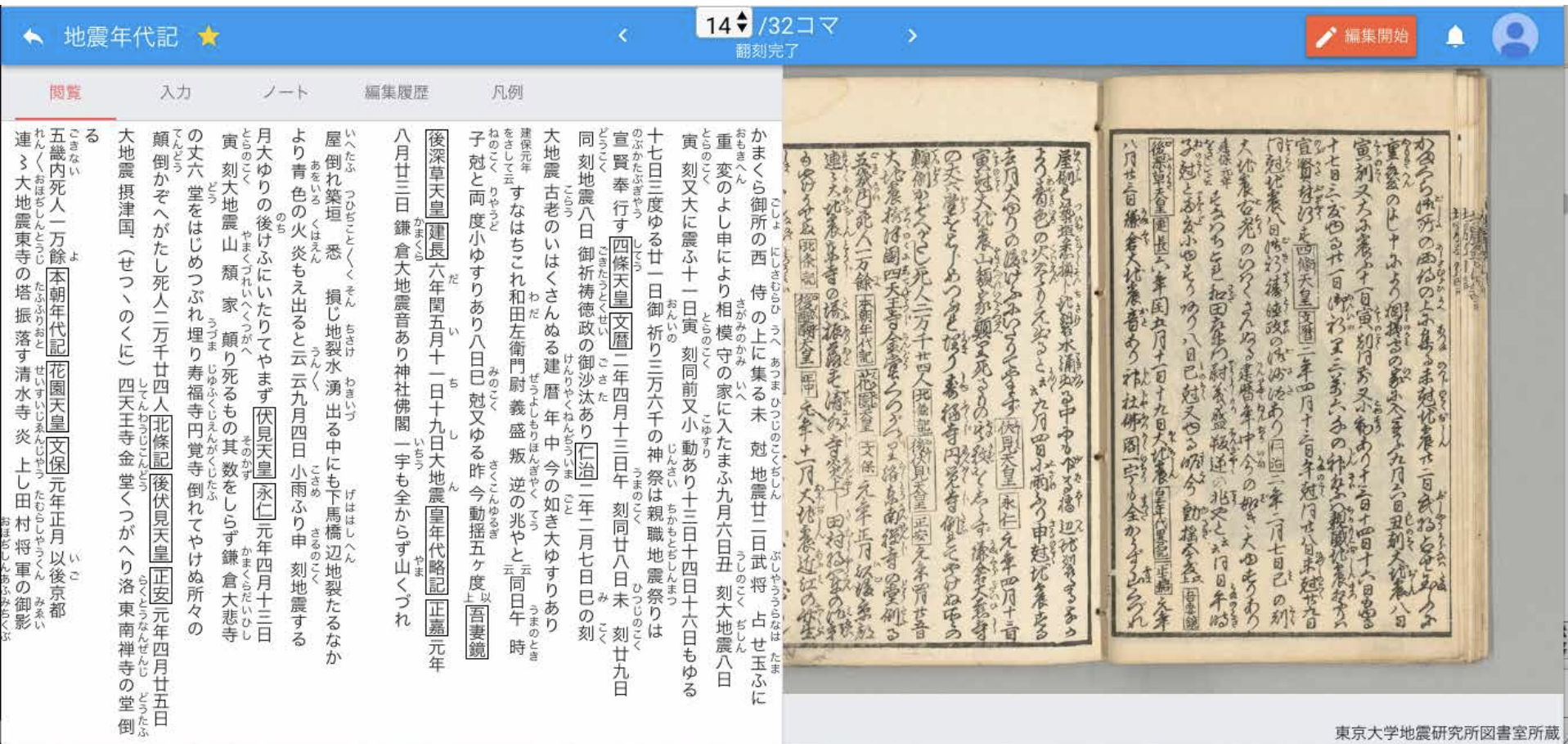
プロジェクトの進捗 (1/10 ~ 2/9)

- Web サイト登録者数・・・1508名
 - このうち177名が実際に翻刻に参加
- 総入力文字数・・・83.9万字
 - 原稿用紙 2200 枚相当
- 画像ごとのステータス
 - 着手済み・・・ 524/3193枚 (全体の16%)
 - 翻刻完了・・・1199/3193枚 (全体の35%)
- 全画像の翻刻が完了した史料・・・35/114点

成果物の品質について

- 非専門家がくずし字を正確に翻刻できるのか？
 - まだ定量的評価は実施していない
- 多数の参加者が編集に関わることで、相当に正確な翻刻ができている印象(集合知的効果)
 - 例:『地震年代記』は、計42人の参加者により、1コマあたり平均12回編集されている

地震年代記(14コマ)



編集履歴

地震年代記 ★		14 / 3
		翻刻完
閲覧	入力	ノート
編集履歴		凡例
すけさん		
2017-01-11 09:01 - 0文字を翻刻		
あっぱれ！		編集内容を確認
すけさん		
2017-01-11 09:01 - 242文字を翻刻		
あっぱれ！		編集内容を確認
つくばもん		
2017-01-11 10:06 - 2文字を翻刻		
あっぱれ！		編集内容を確認
三上正昭		
2017-01-10 12:17 - 1文字を翻刻		
あっぱれ！		編集内容を確認
三上正昭		
2017-01-10 10:38 - 4文字を翻刻		
あっぱれ！		編集内容を確認
三上正昭		
2017-01-10 10:37 - 55文字を翻刻		
あっぱれ！		編集内容を確認

地震年代記 ★		14 / 3
		翻刻完
閲覧	入力	ノート
編集履歴		凡例
seizai		
2017-01-13 06:50 - 626文字を翻刻		
あっぱれ！		編集内容を確認
大河		
2017-01-12 11:09 - 130文字を翻刻		
あっぱれ！		編集内容を確認
十一		
2017-01-12 10:49 - 60文字を翻刻		
あっぱれ！		編集内容を確認
大河		
2017-01-12 09:48 - 12文字を翻刻		
あっぱれ！		編集内容を確認
十一		
2017-01-12 04:19 - 93文字を翻刻		
あっぱれ！		編集内容を確認
steve forrest		
2017-01-12 06:09 - 40文字を翻刻		
あっぱれ！		編集内容を確認

参加者の背景と動機について

- 誰が / どのような動機で参加しているか？
 - 1万字以上入力した参加者が20名もいる
- 後日アンケートやインタビュー調査を実施予定
- 以下では参加者のTwitterでの発言を紹介
 - おすすめ・・・「[みんなで翻刻](#)」で Twitter 検索

Twitter タイムラインを見続けた印象

- 参加者のバックグラウンド
 - 国文学・日本史専攻OBの活躍が目立つ
 - 趣味・教養目的のくずし字学習者も多数
 - リタイアした方が熱心に参加している？
- 参加のモチベーション
 - 「**解読作業そのものの面白さ**」と「**学習意欲**」が継続的参加の強い動機として作用している
 - 「社会貢献」や「学術への貢献」を口にする人は少なめ

まとめ

- 1ヶ月弱の間に、市民の手で史料画像1100枚超、文字数では約84万字をテキスト化することができた
- くずし字解読という難度の高いタスクに対しても、市民参加型の研究は実施可能
 - 地震史料という特殊な題材ではあるが...
- 「学習をベースにしたプロジェクト設計」は、その有用なアプローチになるのでは

次のステップ

- 対象史料の拡大
 - 石本文庫以外の地震研所蔵史料の追加
 - 史料画像のユーザーアップロード
- 別分野への適用
 - 京都学派史料, 台湾植民地時代史料, etc..
 - 国文研オープンデータ？
- 某大規模 Web サービスとの提携
- 機械学習・自動認識の組み込み

書籍『アプリで学ぶくずし字』が刊行されます



- 収録内容
 - KuLAを活用したくずし字教育の実践
 - 海外のくずし字教育
 - 変体仮名文字コードセット
 - 古地震とくずし字解読
 - 刀剣ファン座談会
 - みんなで翻刻の紹介も
- 笠間書院より本日発売
- 92Pフルカラー税別800円