

日本文化とAIシンポジウム2019
AIがくずし字を読む時代がやってきた

Challenges from the past
the future of recognition of scrapped characters
過去からの挑戦状－くずし字認識の未来－



Keio University Institute of Oriental Classics

慶應義塾大学附属研究所斯道文庫

Takahiro Sasaki

佐々木 孝浩

Words of thanks

I would like to thank all those who participated in this Kaguru competition as a researcher of Japanese books written in “Kuzushi ji”.

I would also like to give my heartfelt tribute to **Dr. Tarin Clanuwat, who created the beginning of this competition and symposium.**

慶應義塾大学附属研究所斯道文庫

Keio University Institute of Oriental Classics

- 1960年設立の「日本及び東洋の古典に関する資料の蒐集保管並びにその調査研究を行うこと」(規程)を目的とする研究所。
- 専任教員6名が国内外の東洋古典籍の調査研究を行っている。
- 約17万5千冊(寄託書51,700冊を含む)の蔵書、約6,400本の100フィートマイクロフィルム、約8,100冊の紙焼写真版を収蔵し、専門図書館として研究者に公開している。



文字のない国の苦悩 1

万葉仮名（真仮名）成立まで

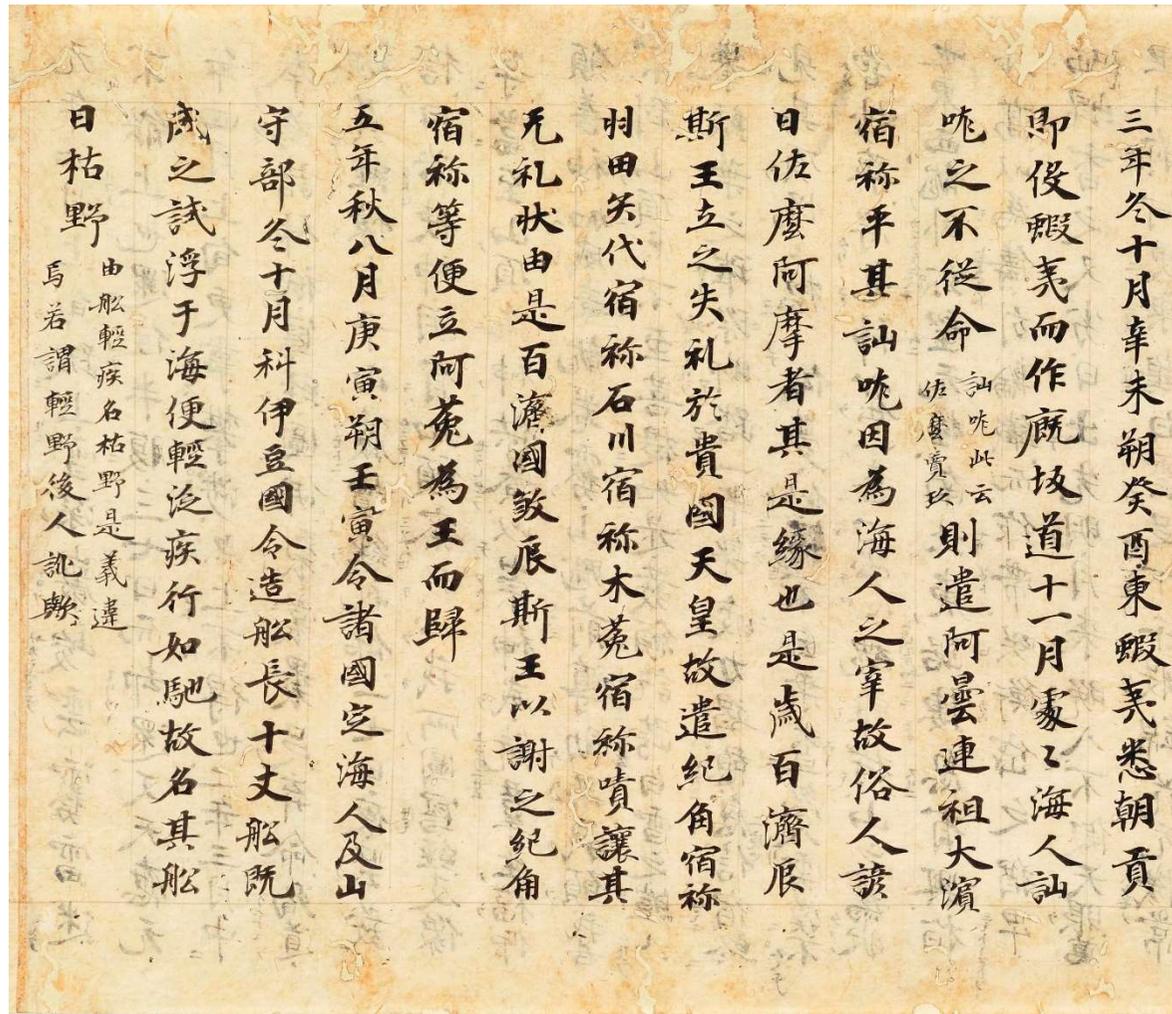
- 日本は固有の文字を持たない国でした。そこに海を越えた大陸の中国から、文字が伝わりました。日本人が「漢字」と呼ぶ中国の文字は、2世紀には日本に伝わっていたとも言われていますが、本格的な伝来は5世紀頃と考えられています。
- それ以後日本人は、漢字を用いて中国語で文章を書き続けることになりましたが、数百年にわたって日本語を記録する方法を模索し続けました。

文字のない国の苦惱 1

万葉仮名（真仮名）成立まで

漢字で書かれた日本人の著作 『日本書紀』養老4年(720年)成立

卷10残卷[9世紀]写(奈良国立博物館)



文字のない国の苦悩 2

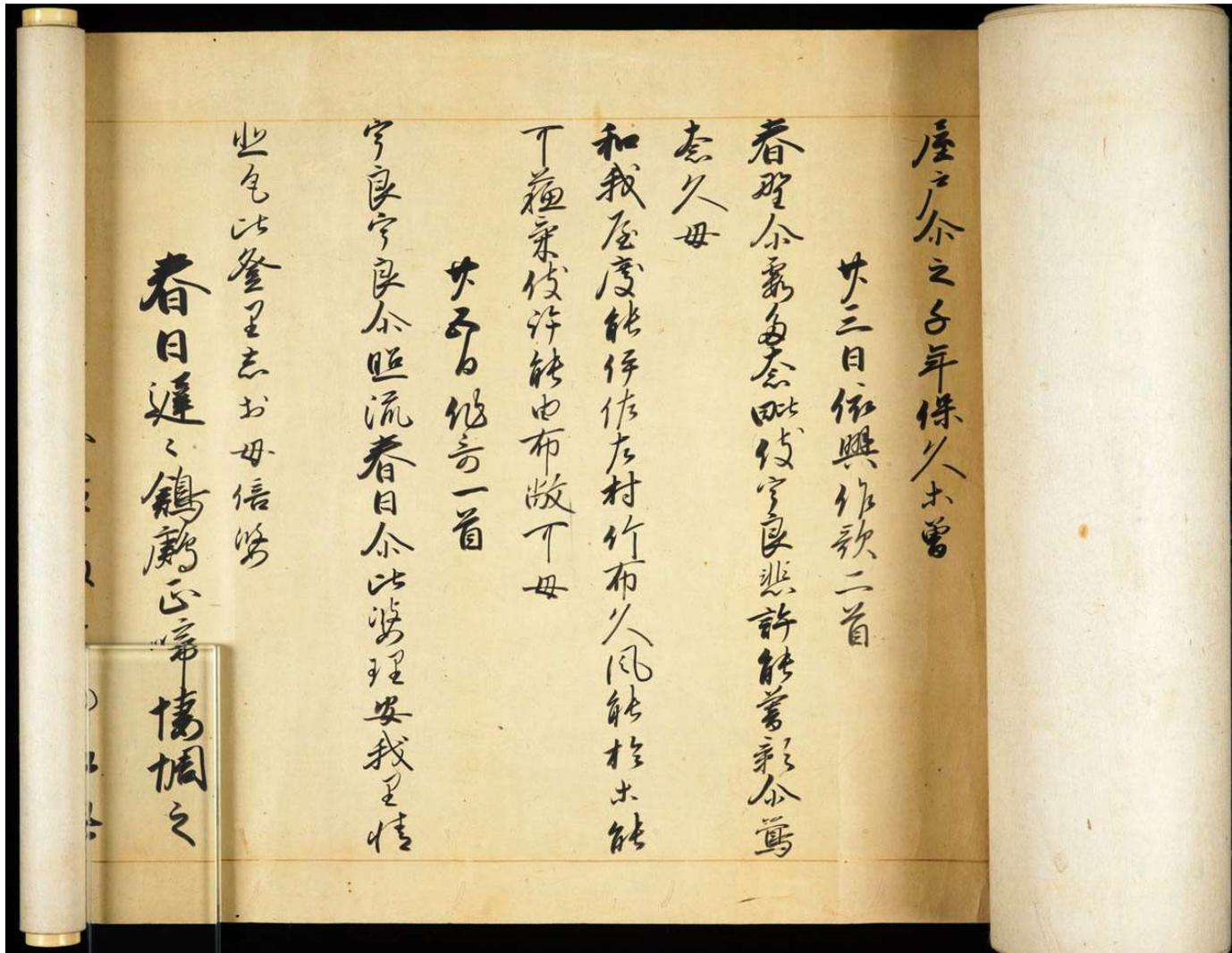
万葉仮名（真仮名）の発案

- 漢字を使って日本語を表記する最初の方法は、表意文字である漢字の意味を無視して、それを表音文字として使用することでした。漢字には中国から伝わってきた発音である「音（おん）」と、漢字の意味に該当する日本語を読み方とした「訓（くん）」があります。この両者を混用しながら、約1000の漢字を使用して、なんとか日本語を文字化したのです。
- この方法は7世紀頃から行われ、8世紀にはかなり発達していたようです。『万葉集（まんようしゅう）』という歌集で本格的に用いられたので、この表記法を「万葉仮名（まんようがな）」と呼びます。「仮名」は正式ではない仮の文字という意味です。正式な文字即ち漢字は「真名（まな）」と呼ばれたので、漢字を使った仮名ということで、万葉仮名には「真仮名（まがな）」の名称もあります。

文字のない国の苦悩 2

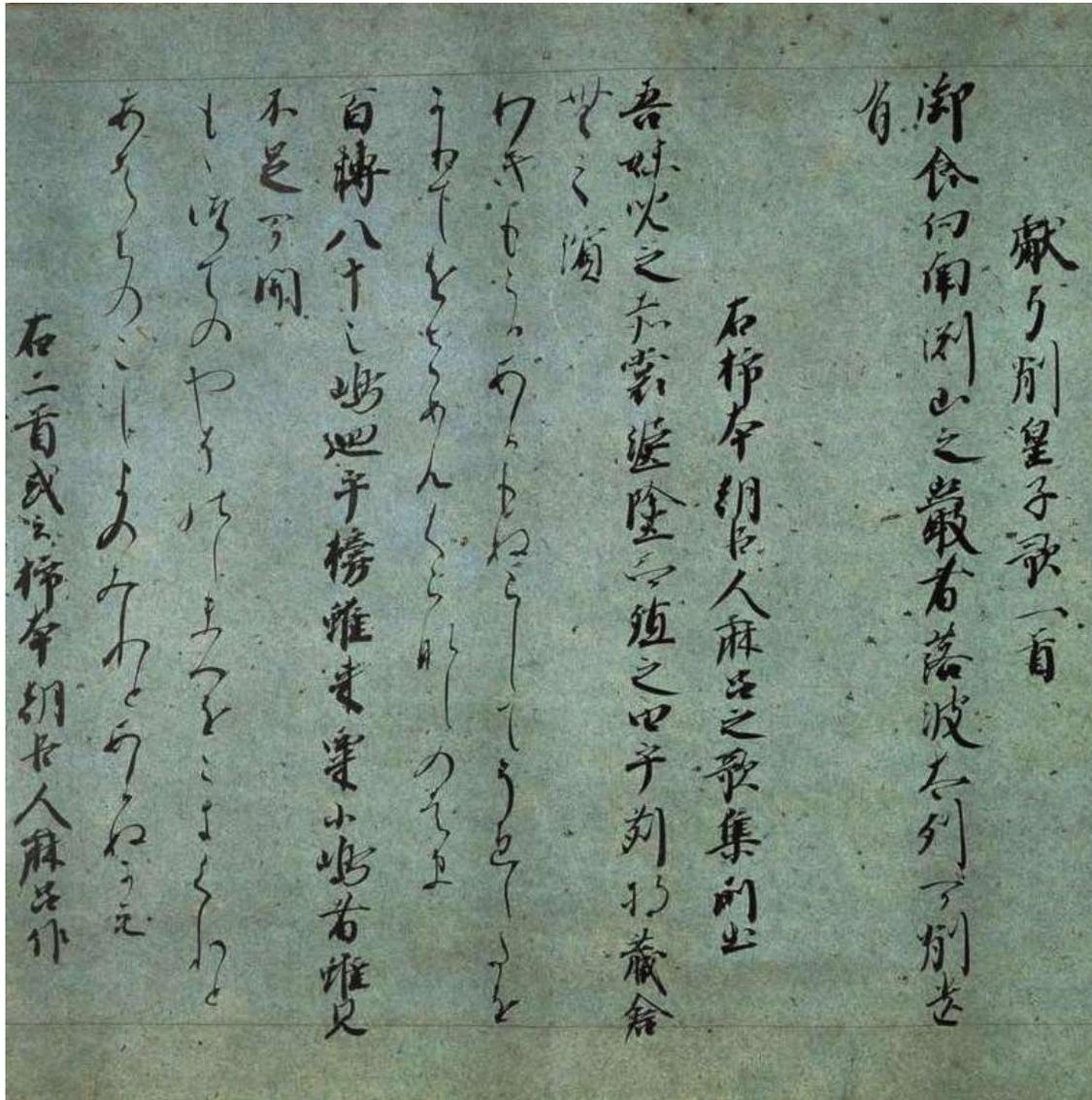
万葉仮名（真仮名）

14世紀の本を17世紀に書写した『万葉集』巻19 理由は不明ながら傍訓を省略して書写しているので、万葉集本来の姿を感じることができる。（斯道文庫蔵）



文字のない国の苦悩 2

万葉仮名と平仮名を並記した『万葉集』



『万葉集』は11世紀の写本断簡が最古のものとされます。それを含めた通常の伝本には、平仮名か片仮名の訓(読み方)が付されている。

『万葉集』平仮名訓本の例

伝藤原公任筆

「藍紙本万葉集」巻9断簡

[11世紀]写

(東京国立博物館蔵)

文字のない国の苦悩 3

草仮名と平仮名

- 同じ意味でも日本語は中国語より文字数が多くなるため、画数の多い漢字を用いる万葉仮名は、書くのにかなりの時間を必要としました。この欠点を補うべく、草書体で書くことが一般化するようになり、草書体の万葉仮名である「草仮名（そうがな）」が生まれることになりました。
- 草仮名をより簡略に書くことから「平仮名」は生まれました。「平」は平易なという意味であるとか、角がない意味などと説明されます。
- 万葉仮名には1文字で数音を表すものもあれば、2文字で1音を表すものもありますが、平仮名では全て1字1音に統一されています。

文字のない国の苦悩 3

万葉仮名から生まれた草仮名（そうがな）

万葉仮名を速筆することから生まれた書体。9世紀中ごろから用いられた。平仮名が完成してからも、芸術的観点や古風さを示すために用いられることがあった。

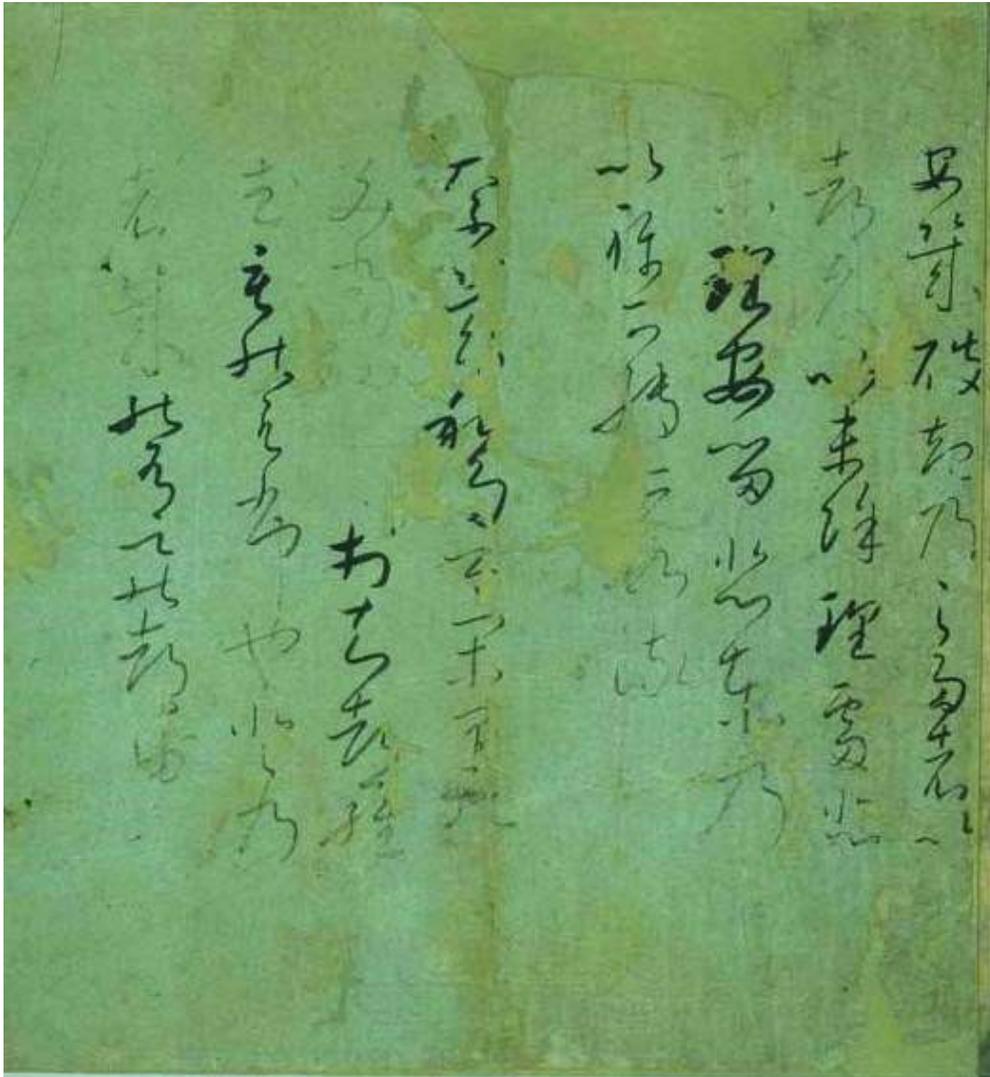
草仮名の代表例

伝小野道風筆

『秋萩帖(あきはぎじょう)』

[10～11世紀写]

(東京国立博物館)



文字のない国の苦悩 4

平仮名

- 昔の平仮名がやっかいなのは、1音に対し多くの文字が併用されたことで、古い時代ほどその数が多く、時代を経るに従ってその数は減少したものの、近代に至るまで複数併用のままでした。
- また判読する上での困難は、同じ漢字から生まれた平仮名であるのに、文字の形にかなりの幅があって一定していないことです。
- 今日の平仮名が基本的に1音1字となったのは、1900年（明治33年）の法令によるもので、そこで選択されなかった平仮名が「変体仮名」と呼ばれるようになりました。

文字のない国の苦悩 4

平仮名

現行の平仮名の「字母」とその草書の一覧

(一覧はウィキペディアより)

无 えん	和 わ	良 ら	也 や	末 ま	波 は	奈 な	太 た	左 さ	加 か	安 あ
	爲 み	利 り		美 み	比 ひ	仁 に	知 ち	之 し	機 き	以 い
		留 る	由 ゆ	武 む	不 ふ	奴 ぬ	川 つ	寸 す	久 く	宇 う
	恵 え	礼 れ		女 め	部 へ	祢 ね	天 て	世 せ	計 け	衣 え
	遠 を	呂 ろ	与 よ	毛 も	保 ほ	乃 の	止 と	曾 そ	己 こ	於 お

文字のない国の苦悩 4

平仮名

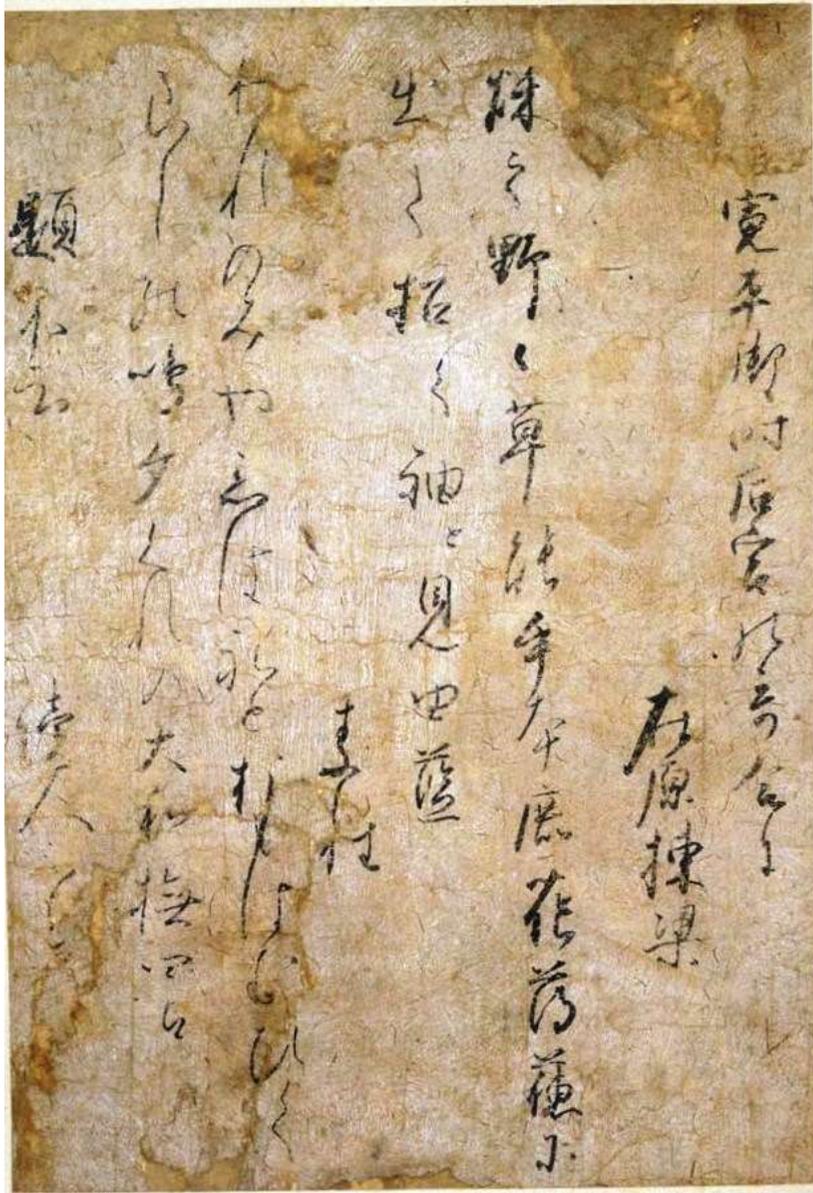
変体仮名の実態「う」の例 「字母」7種

(誠心堂書店HP「変体仮名 五十音順一覧」より)

う					
宇	宇	う	う	宇	宇
有	有		有	有	有
雲	雲		雲	雲	
憂	憂		憂	憂	
そのほか	鶺鴒		羽	羽	右 右

文字のない国の苦悩 5

草仮名と平仮名の混用



- 平仮名が完成しても、草仮名が用いられなくなったわけではなく、書きぶりに変化をもたらすために用いられることもあった。

伝源俊頼筆

「卷子本古今集」断簡

〔12世紀初〕写

(東京国立博物館)

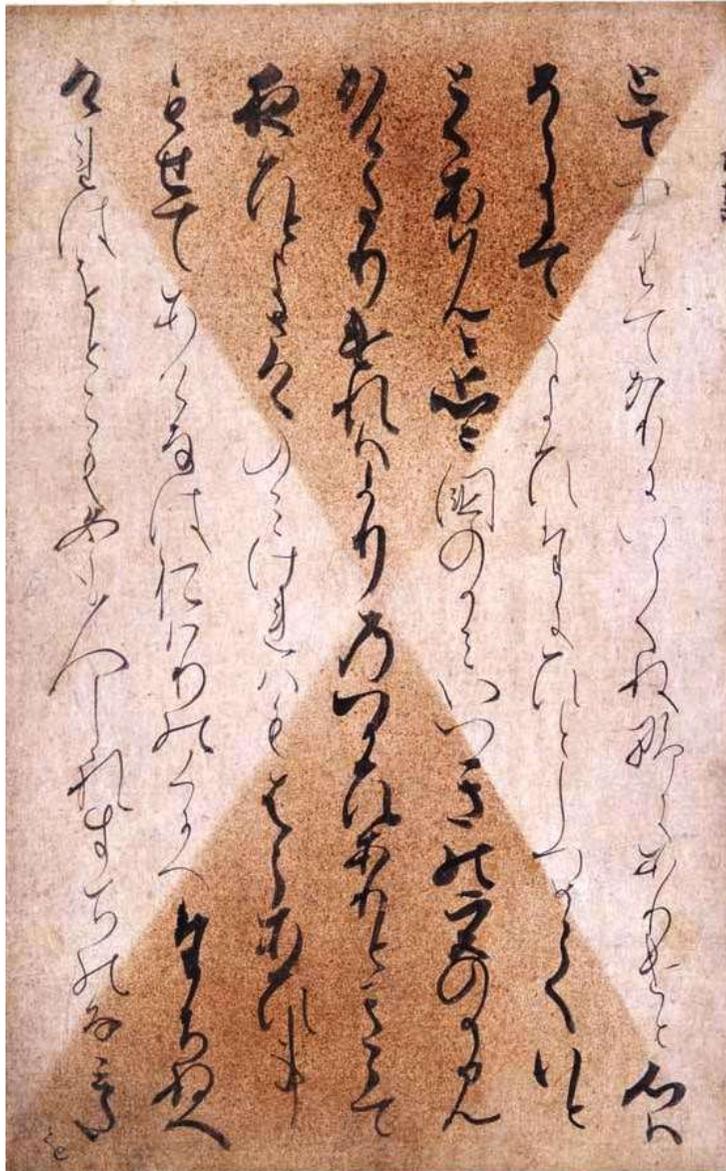
文字のない国の苦悩 6

平仮名とくずし字

- 「くずし字」とは簡単に言うと筆記体のことです。画数の多い文字も線を区切ることなく続けて書きますし、略し方も様々なものがあるので、判読が難しいのです。
- 漢字の総数は膨大で、幾つあるか判らないほどです。一般的に用いられるものだけでも数千字と言われますから、くずし字を読み解くのは大変なのです。
- 平仮名はくずし字から生まれたものであり、基本的に判読が大変ですが、それをより難しくしているのは、平仮名の中に漢字が交じることが多いことです。古い時代は平仮名のみで書かれてましたが、次第に意味を明確にするためなどで漢字の使用率が高くなります。平仮名なのか漢字なのか識別が難しいものもあります。

文字のない国の苦悩 6

平仮名とくずし字



漢字か平仮名か判断が難しいものの例。

漢字でも平仮名の字母でもあり、書く際にそれほど差のないことのある、「葉」・「日」・「三」・「夜」・「屋」等々は、読み進めながら判断する必要がある。

右の例などは「よひとよ」でも間違いではないが、「夜ひとよ」と翻字する人が多いか？

伝藤原為家筆

『伊勢物語』断簡

〔13世紀〕写

(東京国立博物館)

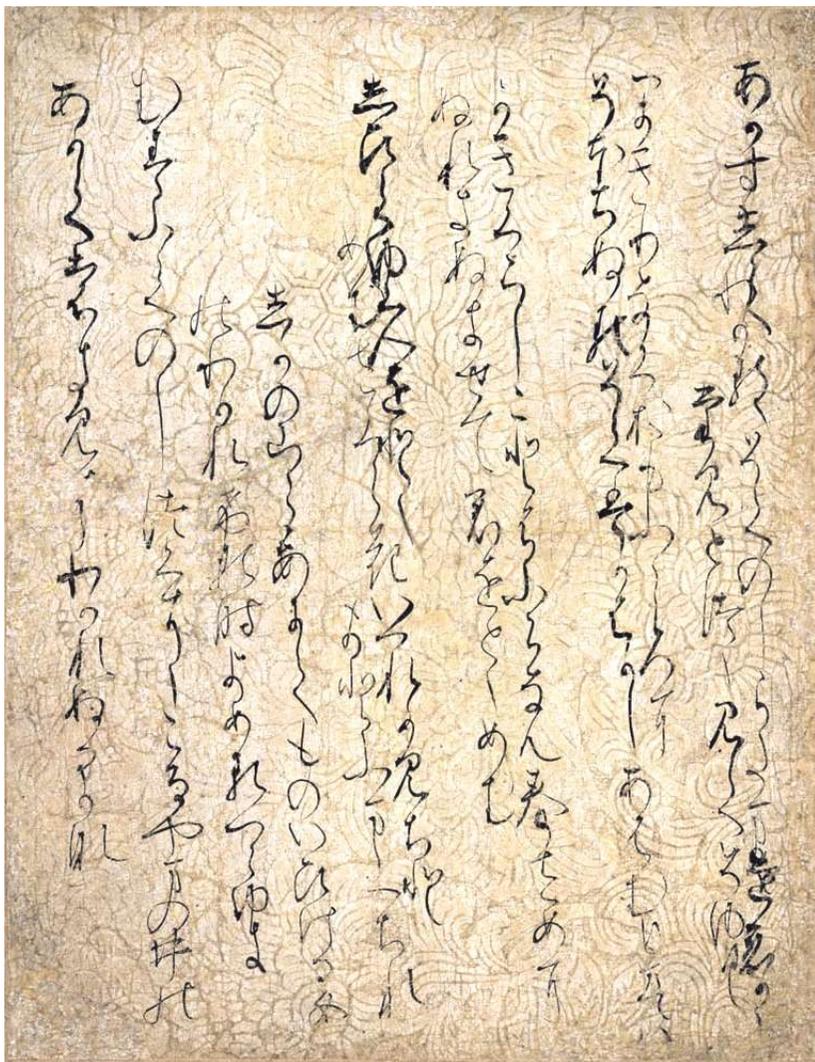
過去からの挑戦

難解なくずし字

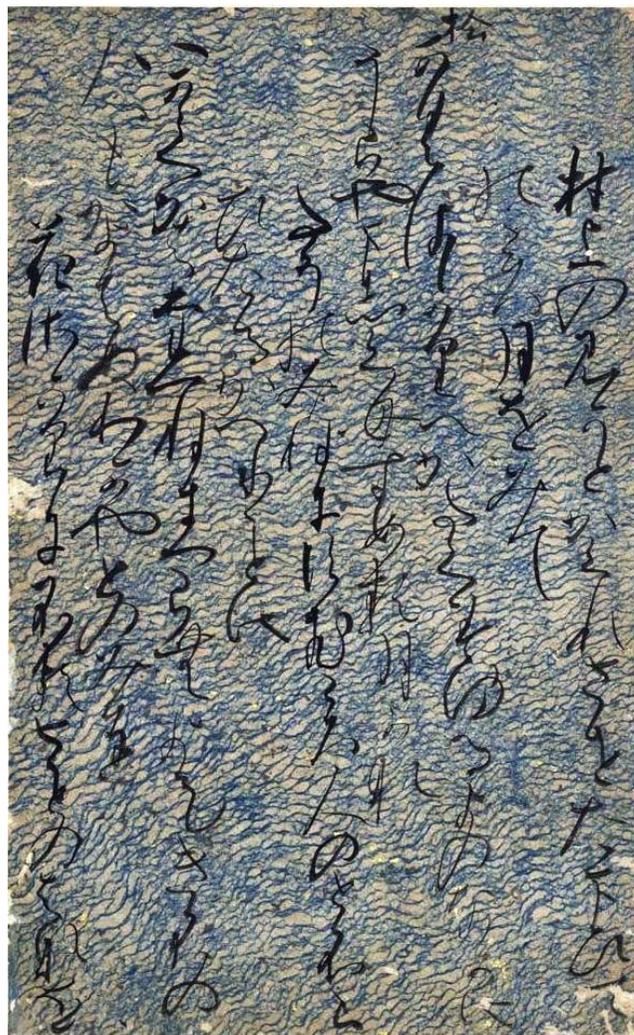
- 平安時代の仮名書き写本には、読まれることを拒んでいるとしか思えないものが多数存在している。その主な理由は以下のようなものである。
 - ① 紙の装飾が文字の判読を邪魔するもの。
 - ② 行がまっすぐではない上に、行間も不均等で、文字が重なり合ったりしている部分がある。
 - ③ 連綿がきつくて文字の切れ目がはっきりせず、非常に曖昧に書かれていて、様々な読み方ができてしまう。

過去からの挑戦 難解なくずし字の例 ①

伝藤原行成筆 「荒木切古今和歌集」
[11世紀]写

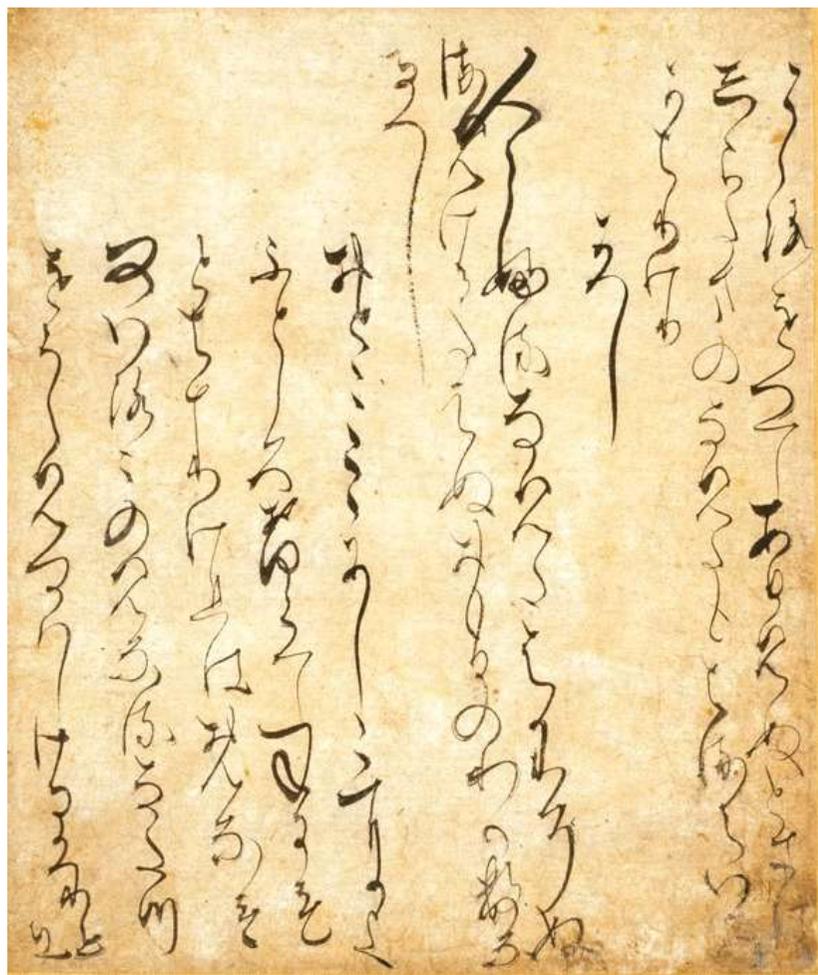


伝源俊頼筆 「高光集切」
[12世紀]写 (共に東博)

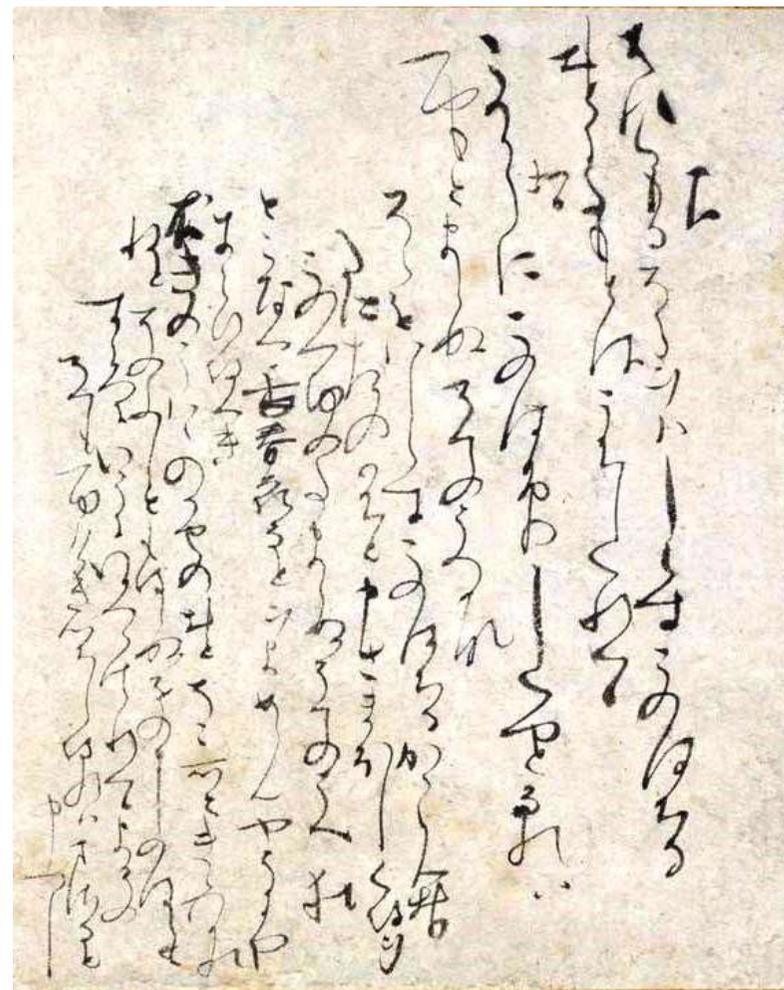


過去からの挑戦 難解なくずし字の例 ②

伝西行筆「白河切後撰和歌集」
〔12世紀後半〕写



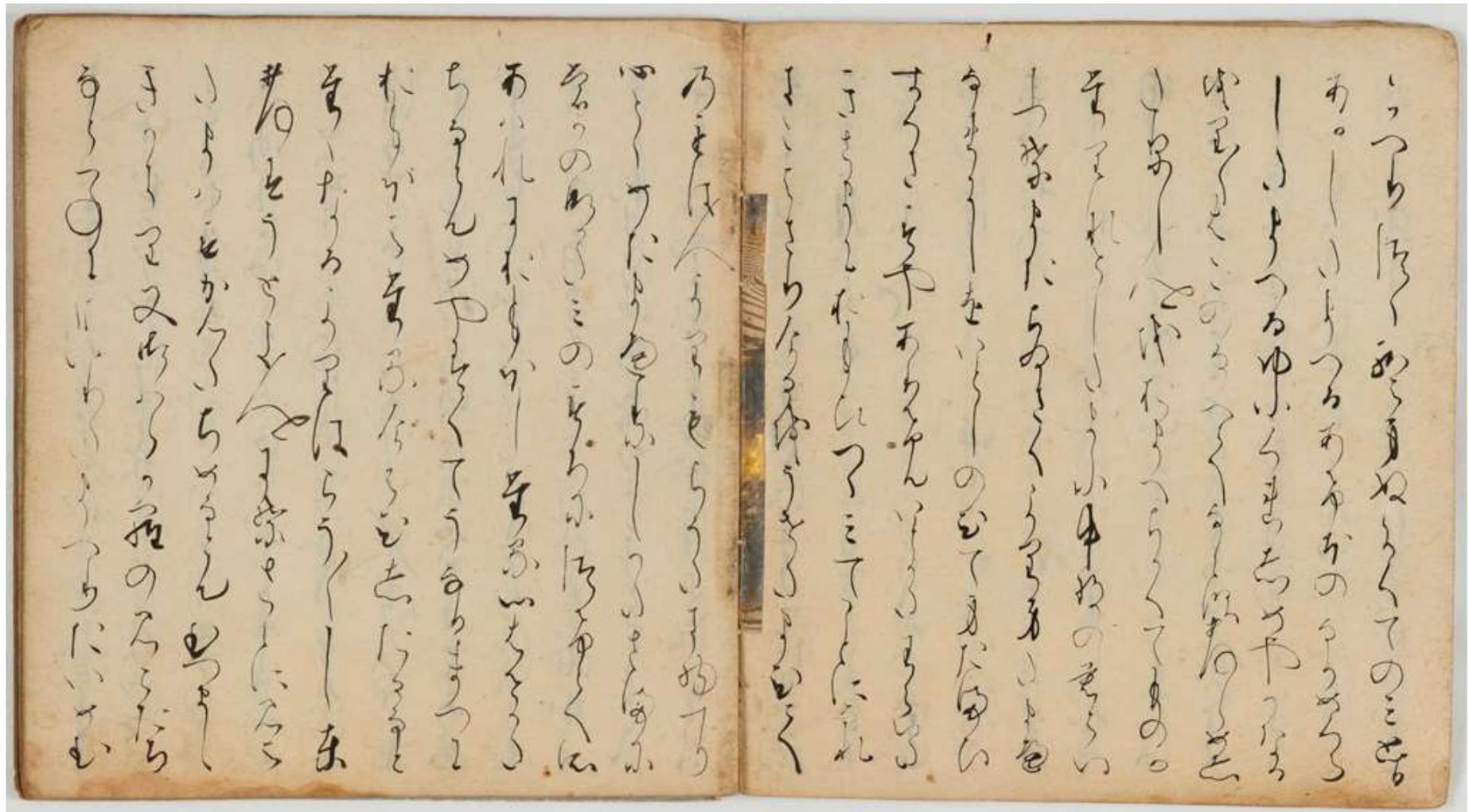
藤原定家筆「歌合切」
〔13世紀初〕写 (共に東博)



過去からの挑戦 難解なくずし字の例 ③

丁寧に書かれていて読みやすいようで実は難読のもの

『源氏物語』 「幻」〔12世紀末～13世紀初〕写 銚子円福寺



くずし字認識の未来 魅惑的な課題 1

- 『万葉集』の万葉仮名は、相当無理をして漢字を当てているところや、まるでクイズのように遊んでいるような箇所も多い。

例 「山上復有山⇒いづ(出)」、「馬声⇒い」、「蜂音⇒ぶ」

- 全歌数約4500首の内、今日でも解読できていない歌が40首程度ある。 ⇒AIで解読できないものか？

卷一・9番歌 幸于紀温泉之時額田王作歌

莫囂円隣之 大相七兄爪謁気 吾瀬子之 射立為兼 五可新何本

仙覚本 ユフヅキノ アフギテトヒシ ワガセコガ イタタセルカネ イツカアハナム

新大観 [] [] わがせこが いたたせりけむ いつかしがもと

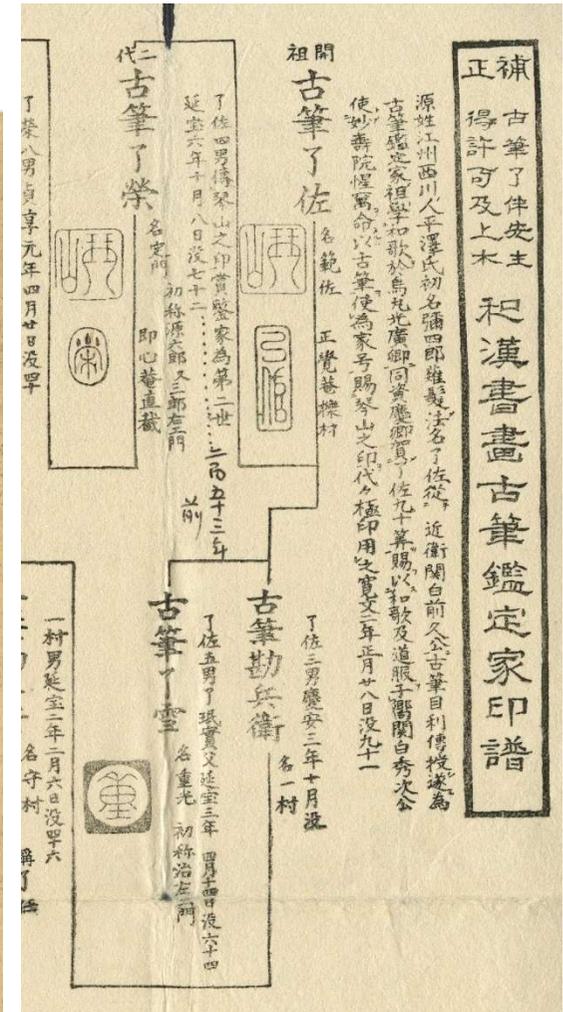
日本の筆跡鑑定 of 歴史

- 日本では17世紀前半から筆跡鑑定を職業とする人物が存在していた。その最初が**古筆了佐**(1572~1662)で、以後その子孫や古筆家の門弟達が鑑定家として近代まで活躍。
- その鑑定結果は相対的に信用できないものであるが、時代の一致率などはかなり高いと評価できる。

公式な文書形式の「折紙 (おりがみ)」、
 小さな紙片の「極札 (きわめふだ)」などの
鑑定書を発行していた。

* → 6代古筆了意の極札 (表と裏)

* ↓ 9代古筆了意の折紙



くずし字認識の未来 魅惑的な課題 2

江戸時代の古筆鑑定家達の仕事をAIで行う

- 書写時代の判定

書風は時代によって変化する。時代の特徴をAIで認識して、描きぶりから書写年代を判定することはできないものか。

- 筆跡の識別 筆者の判定 模写の判別 など

書には流派があって、訓練により文字の形を会得し、師の文字とそっくりに書けるようになる。その筆跡の特徴を把握して、流派名を明らかにするのみではなく、共通性に隠れた個性を見抜いて、筆者の識別や弁別ができないものか？

また価値のあるものには、善意と悪意の模写があるものである。その真贋を識別することは可能なものか？

魅惑的な課題 2

流派や筆跡の識別は可能か？

流派が同じで書写時期が近い事例

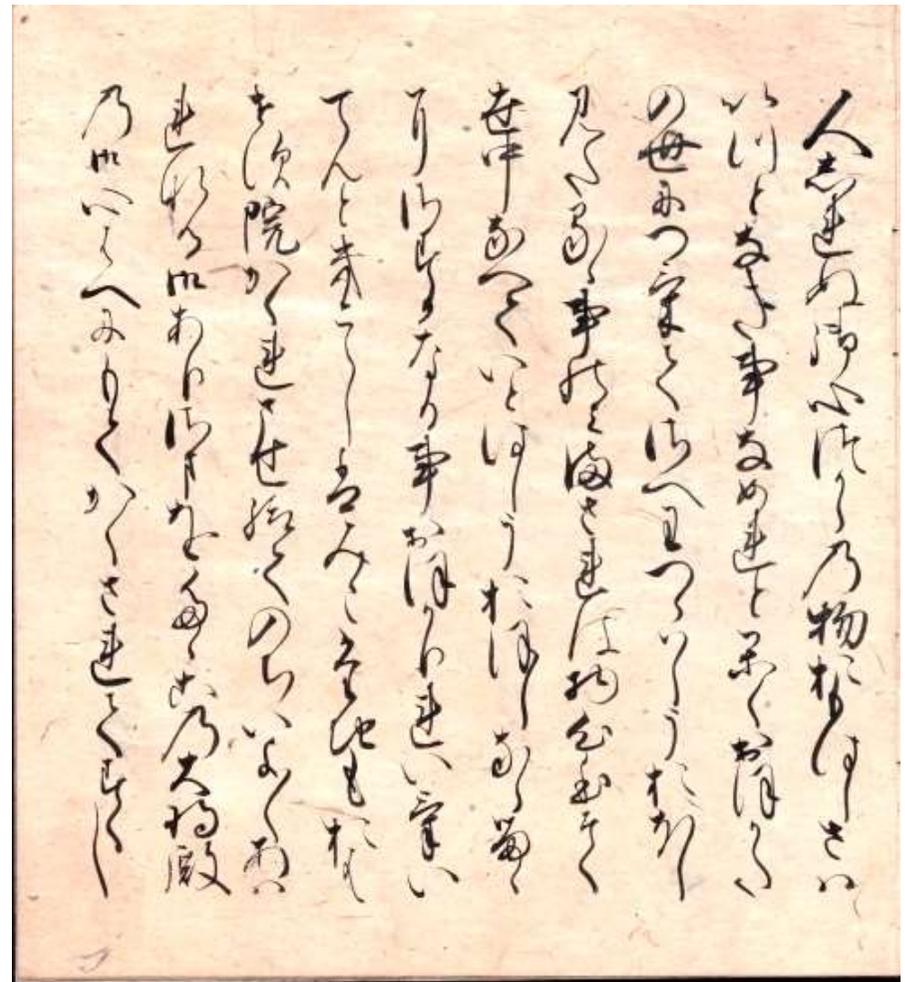
「若菜上」〔16世紀初〕写

三条西公条筆（斯道文庫蔵）



「花散里」〔16世紀〕写

三条西実隆筆（東京国立博物館・保阪本）



同じ流派の人物の文字の比較

上段: 三条西実隆 「花散里」[16世紀]写 (東京国立博物館・保阪本)

下段: 三条西公条 「若菜上」[16世紀初]写 (斯道文庫)

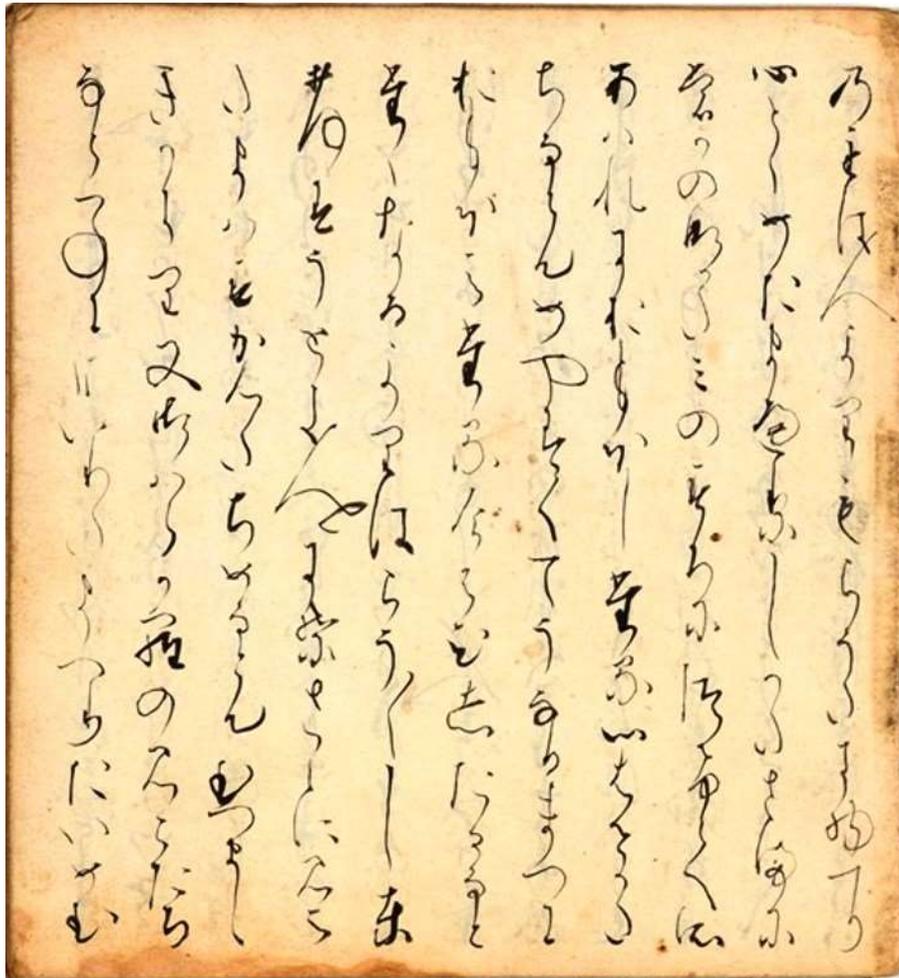


魅惑的な課題 2

書写の前後関係の判定は可能か？

書写年次が近い同じ作品の例

「幻」(銚子円福寺)



「幻」(東京国立博物館・保阪本)



筆跡の異なる同じ作品の同じ位置の文字の比較

上段: 「幻」 (飯沼山圓福寺)

下段: 「幻」 (東京国立博物館蔵 保坂本)

「字母」

梨

裳

東

羅

葉

无

利

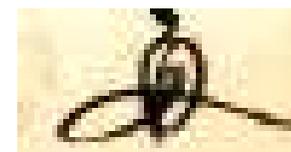
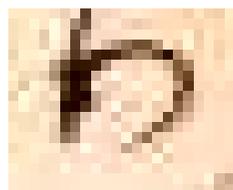
毛

止

良

波

母



くずし字認識の幸福な未来とは

- 革新的な技術の登場は、一般大衆を恐怖に陥れる。人類の歴史はその繰り返しであった。現在進行形のそれがAIなのであろう。
- AIという存在に対しても、日本人の標準的な認識は、「人間の仕事を奪う、なんだか恐ろしいもの」といったところなのではないだろうか。
- 私にAIを語る資格はないが、AIも人類が産み出した道具の一つにすぎず、人が使いこなしていくことが求められるものなのであろうと考える。
- くずし字認識の分野で、今注目されているのは、それを現代の文字に変換する伝統的な「翻字」作業である。それはそれで素晴らしいし、多くのものを翻字していただきたいが、それだけで終わらせてしまっては、技術がもったいない。
- くずし字認識の技術の発達は、これまで日本古典文学や書誌学などの研究者が望みながらも、不可能とされてきたさまざまな事柄を可能にするのではないだろうか。
- それを可能にするためには、様々な分野の専門家の交流を密にする必要があると考えるが、今後さらに必要なのは、タリン博士のような、文理両方の分野を専門とする研究者を養成する環境を整えることではないだろうか。
- 日本古典書誌学の分野は後継者難に喘いでいます。絶対に必要な学問分野なので、これを学びたい若者の登場を願っています。



Thank you very much.

