

アーカイブズの文字 －書字と識字について－

針谷 武志 (史学・文化財学専攻 教授)



1) はじめに

アーカイブズ(永久保存記録)は圧倒的に「手書き史料」である。手書き史料の筆跡や書体はアーカイブズ学でも研究の一分野となっており、手書き史料は、英manuscript, 独Handschrift (あるいは Schriftliche Quellen), 仏manuscritと表記され、また古書体学philologyという場合もある。

今回の話は手書き史料の文字を読むということに焦点をあてる。手書き史料を読むためにもっともよいのは、書けることである。日本の場合は明治以降に活字が普及するが、記録としては手書き史料が作られ続けた。しかも漢字の書体は楷書ではなく草書が用いられてきた。江戸時代にあっては草書が作法とされ、楷書で提出した書類が書き直しを指示された例もある。

日本の活字は上海美華館のW. ガンブルが長崎に活字を移入し、それを種字にして電胎法で母型をつくり活字を鑄造したのが初めのひとつで、本木昌造が継承して長崎の新町活版所、のちに東京と大阪に頒布所がつくられて普及した。美華書館はプロテスタント北米長老会のミッションプレスで、ヘボン『和英語林集成』初版1867年もここで印刷された。中国では明朝体の康熙字典体が用いられており、選択の余地なく日本でもこれが移入された。他の系統の印刷もあり、「楷書活字」も好まれた。明治期は清朝体(弘道軒清朝体)が東京日日新聞に一時用いられた。またいわゆる「教科書体」も昭和5年頃から議論されて昭和8・9年頃から教科書に用いられるが、当時は「楷書活字」と呼ばれた。

康熙字典体(まったく同じではないので「いわゆる康熙字典体」というべきとの言説もあるが)が普及すると、日本の規範字形に変化が生じた。例として、鬱は康熙字典体であるが、日本では鬱が用いられており、草書体も鬱に対応したもののしか用いられていなかったが、前者にとってかわられることになる。

2) 日本は識字率が高いという神話の崩壊

日本は識字率が高いとの神話(根拠なく信じられていること)があるように見

えるが、それを修正する研究がここ15年ほどで多数出されている。先行する実証研究もあったが、ルビンジャー『日本人のリテラシー』（原書2007年、訳本2008年）で注目され、八鍬友広氏が精力的に研究をすすめた。自分の名前が書ける者の割合を「自署率」というがこれは識字率（実用的リテラシー）よりも高い（名前しか書けない者が多数含まれる）。川村肇氏の研究2014で和歌山県の例では識字率は自署率の23%程度と判明しており、八鍬氏が明らかにしたもっとも高い滋賀県の自署率男子87%余も同比率だとすると、高くとも20%程の識字率との暫定的な数値が得られる。欧米の必要とする識字能力はアルファベット20数字なのにくらべて、日本ではおよそ漢字は千字程度の草書体（複数の書体がある）の識字能力が必要ということも考慮すれば、妥当かもしれない。

だが5～20%の識字率でも近世の社会は動いていた、とも言える。渡辺京二『逝きし世の面影』（初版1998年）には文字がよめずとも闇達に暮らしている人びとを観察する外国人の記録が記述されている。

なお「寺子屋」も、明治以降に普及した表現で、江戸時代は「手習塾」「指南所」などと呼ばれたことも定着し、高等学校教科書にも反映している（分析用語としては「寺子屋」はまだ用いられている）。また識字能力獲得の方法は、読み書きではなく、書き読みであったことも主張されおり、手書きという性質上、書き読みの方が適合的である。手書きから活字へ移行するにつれ、読み書きが適合的となる。

3) 日本の手書き史料

日本の実用的リテラシーでは、楷書でなく草書が作法であった。草書にも流派・流儀があり、「御家流」が実用書の基本である。和様（日本風）の一流である青蓮院流の系統であるが、家元制度はきびしく、正規の門弟しか青蓮院流と自称できなかった（田中潤2010の研究がある）。幕府が書流を指定した法令は見いだせず、政策ではなく作法すなわち社会的規範であった。他に三藐院流（近衛信尹の流派、豊臣家右筆など）、嵯峨流（彦根藩岡本家）などもあった。

江戸時代に儒教が普及すると、教養・私的嗜好として唐様（中国風）も好まれるが、実用書は御家流、私的に唐様という二重化が起きた。唐様では趙孟頫、文徵明、董其昌（儒者古賀精里・侗庵らが好む）、顔真卿（佐久間象山、長三洲らが好む）などが好まれたようであるが、残っている揮毫をみるとやはり行草書で楷書はあまりない。

こうした草書の判読であるが、活字なら字形と文脈で読むのに比べて、草書は

字形と筆脈（書き順）と文脈の3つで読む。草書では字形よりも筆脈の方がより大きい要素である。筆で連続で書くと本来ない線が紙に残る。また筆が浮くとあるべき線が紙に残らない。また線がかぶって二重線が一線に見えることもある。字のパーツの大きさや中心線も変化する。筆の動き（書き順）で読むのである。

例として、後深草天皇「宸筆御消息」（1257年以降、田代圭一『人と書と 歴史人の直筆』2016年に掲載）の「事」の字で解説した。筆が浮いて線が欠けて「事」とは判読しにくい、同じ書面に「事」が複数記載されているので、受け取り手は筆脈で読めたであろう。なおこの宸翰は散らし書きで女性に宛てたもので、他の男性にあてた宸翰の「事」とは明らかに字形が異なる。男女でかき分けていたのである。

4) 手書き史料のAI-OCR

日本では多くの研究機関が「史的文字データベース」を構築しているが、多様な文字の書体・書風を反映してバラバラである。奈良文化財研究所「木簡庫」は隸書風の文字、東京大学史料編纂所の「電子くずし字事典」は古文書、国文学研究資料館の「国文研字形検索β」は古典籍の木版本、漢字字体規範史データセット保存会の「HNG単字検索」は漢籍・仏典・国書から集字、などがある。これに台湾中央研究院歴史語言研究所の「簡牘字典」を加えて、奈良文化財研究所が「史的文字データベース連携システム」を構築して横断検索可能となっている。

ためしに、鬱と鬱を検索してみると、東大史料は鬱の草書体を見出し「鬱」と「鬱」の両方に登録している。実際の字形で「鬱」は一字もない。異体字を区別していないということであろう。国文研は鬱の行草書を見出し「鬱」にのみ登録して、異体字を区別している（やはり実際の字形は「鬱」は一字もない）。全体としても実際の字形「鬱」はHNGの2例あるのみだが、漢籍で中国での使用である。HNGには実際の字形に「鬱」も7例もあるのだが、すべて見出し「鬱」に登録して、見出し「鬱」に登録していない。異体字を区別しない上、実際の字形「鬱」も見出し「鬱」にしてしまっている。もっとも厳密なのが国文研、異体字研究に支障あるのがHNGということになる。

さてAI-OCRは、いくつかある様であるが、TOPPAN（凸版印刷）の「ふみのは」をとりあえず見てみると、三井文庫の史料からくずし字データベースをつくって教師データ（訓練データ）としてAIに学習させたものと記載されている（三井文庫「AIを用いたくずし字解読と三井文庫の取り組み」チラシ）。これをスマートフォンのアプリに移植したのが「古文書カメラ」で、web上にはこれを試した実

例が多数いろいろ掲載されているが、解読率0～80%程度と偏差がはげしい。多様な文字の書体・書風からすればむしろ当然とも思われる。AIには目だけで手が無いので、書き読みができないことが根本的問題と思われる。訓練データに書き順を紐付けすれば多少は向上するだろう（紐付けしないと限界を超えられない）と思われるが。

5) 「限界リテラシー」

多重な文字・言語・文体が用いられ、多重なりテラシーがあることは西洋史研究でも指摘されている。大黒俊二氏の研究(2010, 2014, 2019)では、中世イタリアでは俗語とラテン語の二重りテラシーがあり、1427年フィレンツェの俗語識字率（実用的りテラシー）は成年男子の69.3%で、代筆の事情によってはさらに上回るだろう、としている（『声と文字』2010）。日本の20%という暫定値よりはるかに高いことに驚く（日本では子どもも含んでいることにも留意する必要あるが）。また「限界リテラシー」の指摘は示唆ぶかい。これは、読めるが書くことに問題がある、境界線上の人々のリテラシーで、個人レベルで勝手な文字、文法をつくる現象が起こる。たいていは追い詰められた状態で起こり、端的な例は魔女裁判での自白書（エゴ・ドキュメント）でむしろ事実ではない内容という。

日本でも読解にくるしむ、勝手な造字ではないかと思われる文字に遭遇することがある。すくなくとも把握されている異体字ではない。追い詰められるという状況では必ずしもないが、限界リテラシーに近い現象ではないかとの可能性を考えている。もしそうであるなら、勝手な造字が行われる状況下で、最初の戸籍である明治5年(1872)壬申戸籍（明治5年式戸籍）がつくられたのかもしれない。現在の戸籍法と法務省令では戸籍通りの文字を書くことを要求しているが、これは悪手ではなかろうかとも思うのである。

6) おわりに

今回は手書き史料、それも実用的りテラシーに焦点をあてたが、大黒氏もほかに「文学的りテラシー」に言及している。これは実用上必要だからではなく、文学を楽しみたいから識字能力を得たいという希求が人々にあるということであり、日本でもとくに女性の識字能力希求に関わるのではないかとも思うが（女性俳人の田上菊舎や樋口一葉などが思い浮かぶ）、今後の課題（楽しみ）としたい。