画像デジタルアーカイブの文字研究への利用

――「祇園」の表記研究とメタデータの問題点――

當山 日出夫（立命館大学文学部・非常勤講師）
E-mail：htt04112@pl.ritsumei.ac.jp

要旨：京都の祇園の表記史研究において、立命館大学アート・リサーチセンター（以下、本稿ではARCと略称する）のデジタルアーカイブがきわめて重要な研究資料となる。これの利用によって、祇園の祇について、右旁（つくり）を祇と書く事例は、単なる誤字ではなく、京都の伝統的標準慣用字体であったことが判明する。デジタルアーカイブを実際に使った研究成果の報告である。また、デジタルアーカイブの有効利用のためには、検索メタデータの適切な付加が不可欠である。本稿は、この視点からの報告をかねるものである。

1．はじめに―祇園の表記の問題点―

筆者は、昨年来、京都の「祇園」の表記の実態をめぐって、種々の調査研究をおこなっている。文字史の観点から、これまでに判明したこと、および、論点を整理する。以下に述べることは、狭義では文字論・文字コードの問題ではある。しかし、後述のメタデータの記述とも関連する課題であるので、概略を記す。

(1).「祇園」の「祇」の文字は、旧来の83JIS漢字（JIS X 0208）では、「しめすへん＝ネ」・「つくり＝氏」の字体であり、一般のワープロ文書やコンピュータ表示では、この字体が使用されている。しかし、マイクロソフト社製の新OS（WindowsVista）においては、新しい04JIS規格（JIS X 0213:2004）が、標準として採用される[(1)]。この場合、「祇」については「しめすへん＝示」の字体に変更になる。これは、さらにさかのぼって83JIS以前の78JISの字体に戻ることでもある。つまり、「祇」は、コンピュータで使用する文字としては、次のような変遷をたどることになる文字である。表記（印刷）の便宜上、「へん」と「つくり」を「ネ」「示」「氏」で表示する。

78JIS（示氏）→83JIS（ネ氏）→04JIS（示氏）

(2).「祇」は、現実的な使用例としては「祇園」の地名表記に集中的に使用される[(2)]。では、実際の京都の祇園地区においては、この文字はどのように表記されているであろうか。この問題点について、全国的に均一な文献（活字）資料ではなく、京都の祇園の現地における表記について調査した結果、次のような結論と課題を得るにいたっている。

①資料としては、文献資料（印刷・活字）ではなく、非文献資料を主として対象とする。非文献資料とは、道路標識・バス停・観光案内図・店舗の看板・地名表示案内など、まさに祇園の現地で実際に目にする文字である。

②「しめすへん」について観察するかぎり、京都の伝統的花街としての風情をたもつ地域では、基本の字体が「祇（ネ）」である。一方で、現代的な繁華街としての地域においては多様な「祇園」の表記がみられ「祇（示）」も多く使われる。また、京都市バスについては、従来「祇（ネ）」であったものが「祇（示）」に変更されていることが確認できる。つまり、京都の祇園の基本的な表記は「祇（ネ）」である。

(3)．一方、「つくり」について見ると、「しめすへん」がいずれであろうと、基本は「氏」である。しかし、そうはいうものの、少なからざる異例が見いだせる。

①「氏」の下に「一」、右肩に「てん」を付加する字体である。このうち、「氏＋一」の字

体は、漢字としては「祇」（※「つくり」にのみ注目してほしい）となり、「祇」と「祇」は、現在の通行の漢字字典の規範によれば、別の漢字である。

②「しめすへん」を「ネ」で書くか「示」で書くかについてほど、使用例や地理的分布からの分析が可能な程の事例を、採取できない。つまり、「氏＋一・てん」の字体が、単なる誤字であるのか、あるいは、伝統的な祇園の表記を残すものであるのか、歴史的に検証の必要がある。

ここで、文字の歴史的研究におけるデジタルアーカイブの利用とその問題点が課題となる。以下、本稿では、京都の祇園の表記研究を通じて得た私見の一部を整理し、述べる。

2．祇と祇の表記の実例

まず、図（1）（2）に現在の祇園における事例図（3）（4）にARCのデジタルアーカイブにみる近世の事例[4]をしめす。ここで使用したのは、ARCの「書物と絵画プロジェクト／京都名所図閲覧システム」である。

紙数の都合上、限定的な事例しか提示できないが、その気になって探せば、すぐに見つかるといってもよい。かなりの「祇」の事例が、京都の祇園において観察される。そして、これは、結論としては、近世期以来の京都の伝統的な祇園の表記をうけついでいるものと判断できるのである。

3．デジタルアーカイブ画像の価値と課題

ここに述べたような結論が得られたのは、まさにデジタルアーカイブの存在によってであることは言うまでもない。従来であれば、図書館・博物館などの目録を検索して、逐一実物を見なければならなかった。また、浮世絵類は、保存のため貴重書あつかいであるので、その閲覧手続きにかなり労をついやさなければならかった。しかし、デジタルアーカイブの出現により、非常に研究が容易になったことは確かである

図（1）

図（2）

図（3）

図（4）

しかし、単にデジタルアーカイブを評価するというわけにはいかない。ARCのデジタルアーカイブは、今回の筆者の研究には、きわめて有効に利用することができた。だが、その過程で、以下に述べるようないくつかの問題点を感じるところがあったので、一般論をふくめて、忌憚なく述べることとしたい。

4．画像の精度

検索・一覧のためのサイズの小さいサムネイル画像は必要である。その一方で、書いてある文字が確実に判読できるだけの高精細画像の提供も必要である。デジタルアーカイブは、単に画像データを見るだけのものではない。現在では、実物の保存のため、その閲覧のかわりを果たすべき役割をになっている。したがって、書物・文書類はもちろんであるが、絵画類であっても、中に文字の記載のあるものについては、それが十分に判読可

能なレベルの画像が求められる。

したがって、一律に、画像のサイズ（ピクセル数）を決めてしまうのではなく、臨機応変に対処する方向があってよい。部分的に細かな文字が記されているような画像については、その部分だけの拡大画像が、別に用意してあるのが望ましい。もちろん、絵画研究の立場からしても、描写の技法などの精査のためには、可能な限りの高精細画像が要求されることはいうまでもない。ARCのような画像デジタルアーカイブであっても、筆者のように、そこに「文字」を探す研究者もいる、その研究者の目的の多様性に対応できることが必要である、ということである。

画像精度の点については、ARCは、筆者としては十分に満足できるデータを提供してくれている。

5．メタデータ

今回、筆者が、対象としたのは、「祇園」という文字である。そのため、検索においては、江戸時代における現在の八坂神社（旧称は「祇園社」）と、その祭礼行事である祇園祭を探せばよい、という比較的容易な状況であった。したがって、検索のキーワードは、と「祇園」「八坂神社」が基本になる。

実際にARCで検索した場合、「祇園」の検索結果と、「八坂神社」の検索結果には、かなりの違いがある。次のような例である。

(1).「祇園」が概念として最も広い。「祇園社」としての名称の他に、「祇園祭」とそれに関係するもろもろのデータが検索対象となる。

(2).「八坂神社」の場合、まさに神社としての八坂神社に限定されてしまう。

(3).「八坂神社」で検索した結果、表示されるのは、「祇園社」であり「八坂神社」の文字が見あたらないものもある。

(4).「祇園祭」と「祇園会」の検索結果は同じではない。

おそらく、これは、検索のためのメタデータを付加した人間の判断の違い、ということによるのであろうと推測される。確かに、旧称は「祇園社」であり、現在の名称は「八坂神社」である。この場合、歴史的な資料として、どちらの名称を選択すべきか、判断が分かれるのは、十分に納得できる事情ではある。

このような場合、次のような各種の対応が考えられる。

(1). この種の学術的デジタルアーカイブは、利用者が、ある程度の専門知識を持っていることを前提としておく。したがって、「祇園社＝八坂神社」であることを知っており、両方を検索すればよいとして、その判断は利用者にゆだねる。かならずしも、オリジナル資料に忠実でなくても、かまわない。

(2). あくまでもオリジナル資料の記載に忠実である。「祇園社」とあれば「祇園社」で検索でき、「八坂神社」とあれば「八坂神社」で検索できるようにしておく。この方式の場合、「祇園社＝八坂神社」であることはかなり知られていることなのであまり問題は生じない。しかし、このとき、本稿で問題とした「祇園」（現在の通行の表記からは誤字とされる）を、どうあつかつかうか、問題となる。

(3). 現在の通行の地域名称・施設名称と、オリジナル資料での記載とは、別扱いにする。この場合でも、現在の通行の名称として何を採用するか、また、オリジナル資料の記載から何を採用するか、かなり難しい判断を要求されることになる。

今回の筆者の「祇園」研究での利用によっても、以上のような問題点が浮かび上がってくる。これには、さまざまに対応があり得る。筆者の研究目的（＝文字史・表記史の資料）という観点からは、次のように考える。

(1). メタデータは、多目的に対応できること。たとえそれが、画像デジタルアーカイブとして浮世絵であっても、あるいは、陶磁器であっても、その中に文字がある限り、文字研究者にとっては、資料になる。文化財の安全な保管とい

う意味でのデジタルアーカイブ化を進行する以上、実物を手にとって見るに匹敵するデータの提供が望ましい。そして、その利用目的（研究対象）は多様である。

(2)．検索キーワード（メタデータ）はあまり限定的に絞り込まないで、ある程度、おおざっぱな方が、かえって使いやすい。この点は、たぶん、理工系（情報学）と人文系の研究者の、発想の基本にかかわる点である。あえて強引に一般的に述べれば、

①情報学系研究者であれば、あるキーワードに対する検索ヒット率の向上が目的・目標になる。したがって、キーワードを細分化する方向に向かいがちである。（しかし、これは、研究者の関心の多様性に対応が難しいという難点をかかえる危険がある。）

②しかし、人文学系研究者は、資料をしらみつぶしに自分の目で見て探す、という作業を、さほど苦にはしない。あるいは、むしろその過程で出会う種々の副産物的な発見の方に、より興味があるといってもよい。（とはいえ、あまり厳格に絞り込まれたキーワードの範囲を超えた調査のため、当該デジタルアーカイブの全データを網羅的に見なければならないというのは、効率的とはいえない。）

6．研究者からのフィードバック

今後、デジタルアーカイブを有効に活用するためには、さらに次のような課題がある。

(1)．デジタルアーカイブのメタデータがどうあるべきか、についての、利用者（研究者側）と作成者の対話が重要である。かりに、文字資料に限定しても、言語研究の立場からの文字史と、美術史・書道史研究の立場からでは、微妙に異なる。現実には、そこにデジタルアーカイブ作成者の意図がどうであるかが、からんでくる。多様な立場からの学際的な意見交流の場が必要である。

(3)．メタデータの付加、場合によっては正誤の訂正などは、研究者の高度な学術的判断が要求される。その場合、ユーザ（多様な分野の専門研究者）の意見やコメントをなにがしかの方法で反映するシステムは必要である。しかし、ただそれを、そのまま加えればよい、というものではない。そこには、デジタルアーカイブの作成・管理者の、各専門分野の研究者に対応するための、検証と選定の作業が、必要になってくる。

(4)．人文学系研究者の専門の世界はきわめて狭い。そしてそこには複雑な人間の関係がある。このなかで、純粋にアカデミックに学術的判断の認定をめぐって議論を蓄積していくことが可能であろうか。この点について筆者は、長期的に将来的な展望としては決して悲観的ではない。しかし、短期的に見れば、最悪の場合、かりにミスが有ったとしても、ミスのまま放置されてしまうということがあり得ないではない。研究者個人レベルで、各自の学術的判断にしたがって、補正すればすむことのなのである。この点が、さしあたっての最大の課題かもしれないと考える。

(5)．いうまでもなく、学術資料のデジタルアーカイブの構築と利用をめぐる、現在の緊急の課題は、相互の総合的な横断検索である。幸い、ARCは、この方向性については先見の明があるシステムとしてきわめて高く評価できるものである。しかし、さらに他の多くの研究機関などで作成のデジタルアーカイブとともに統合的に検索できるシステム（そして、それは、各種の研究者の研究目的にかなった柔軟なものである必要がある）の構築を期待したい。そのためには、利用者（研究者）からのフィードバックが重要な課題となる。

7．おわりに―今後の課題―

各種学術資料の、国家規模でのデジタル化が、昨今の最重要課題であることはいうまでもない。しかし、その実現は我が国においてはそう容易で

はない。困難は多い。だが、そのためには、「デジタルアーカイブ化すべき文化資産とは何か」「その利用の問題点は何か」についての各分野の研究者間における、率直な対話が必要である。本稿は、ささやかながらも、この視点からの提言をふくむものとしたい次第である。

付記

本稿は、『第12回公開シンポジウム　人文科学とデータベース』（2006年12月23日、於大阪大学中之島センター）での口頭発表「祇園を祇園と書くのは誤字か―浮世絵アーカイブを資料とした検証―」のうち、特にメタデータについて述べた部分をもとに、再構成・加筆したものである。なお、文字史・文字論については、別の機会での学会発表・論文執筆を予定している。

また、本稿は、2006年12月の時点で執筆しており、実際に本誌が刊行されるときには、WindowsVistaは既に世にでているという状況である。文字コードの規格と実装についての把握に不備な点があるかもしれないが、御寛恕をこう次第である。

注

(1)．04JISの名称は、現在では、普及したものではない。本来ならば、規格票の名称として、「0208」と「0213」は区別すべきである。しかし、マイクソフト社のHPにおいても、新フォント（メイリオ）の解説において、「JIS2004」と称している。
http://www.microsoft.com/japan/windowsvista/jp_font/default.mspx
JIS X 0213:2004 対応と新日本語フォント「メイリオ」について

(2)．「祇園」以外の使用例としては、「神祇」あるいは「祇王（寺）」などがあるが、そう多く観察できるものではない。

(3)．図（1）は、花見小路（南）の飲食店の街頭に掲示のメニュー。図（2）は、北側の神社の鳥居。

(4)．図（3）はUP0701、図（4）はUP1828。

参考文献

赤間亮（2005），「古典芸能研究におけるデジタルアーカイブの効用」，『第11回 公開シンポジウム 人文科学とデータベース』，（2005年12月 3 日，於大阪樟蔭女子大学）

當山日出夫（2006a），「京都における「祇園」の表記の実態―非文献資料による文字史のこころみ―」，第94回訓点語学，（2006年 5 月12日，於東京学芸大学）

當山日出夫（2006b），「京都における「葛」と「祇」の使用実例と「JIS X 0213:2004」―非文献資料に基づく考察―」，『情報処理学会研究報告』（2006-CH-70）

當山日出夫（2006c），「地名用字の今むかし―京都の「祇園」の場合―」、『日本語学』（2006年12月号），明治書院

當山日出夫（2007a），「非文献資料による文字論の課題―京都における「祇園」を主な事例として―」，『立命館大学白川静記念東洋文字文化研究所紀要』（第 1 号）

當山日出夫（2007b），「京都の「祇園」の表記―「しめすへん」をどう書くか―」，『国語文字史の研究』（第10号），和泉書院，

『文化情報学のパースペクティブ―デジタルアーカイブへの新地平―』（人文科学とコンピュータシンポジウム「じんもんこん2006」論文集），情報処理学会，（2006年12月14・15日，於同志社大学）