

古代木簡研究におけるデジタルデータの整理と集積

井上幸^{†1}

古代木簡研究において重要視されるさまざまな情報について、実際の木簡解読作業を例に述べる。また、これらの情報を情報技術を用いて整理・集積する際の注意点と工夫について紹介する。

Digital data arrangement and accumulation on the study of the ancient Japanese wooden tablets

MIYUKI INOUE^{†1}

I describe various information regarded as important in the ancient Japanese wooden tablets' study, giving an example of real decoding work of wooden tablets. I introduce matters that require attention and device to arrange and accumulate it using information technology.

1. はじめに

発掘された木簡の記録は、1点1点、観察記録や写真という形で蓄積されている[a]。この記録は、脆弱で容易に出し入れのきかない木簡にとってかわる記録であり、後々まで解読や研究に不可欠な、重要なはたらきを担っている。

その記録作業の方法は、長年のノウハウが引き継がれ、今に至る。それと同時に、昨今のデジタル技術のめざましい進展によって、これらの作業を取り巻く環境が変わりつつある。これらの技術を利用して、これまで紙やフィルムなどのアナログな媒体によって集積してきたデータのデジタル化も進めている。

これらの資源の活用については、2003～2007年度に採択された科学研究費補助金基盤(S)「推論機能を有する木簡など出土文字資料の文字自動認識システムの開発」(課題番号15102001)に始まり、以降、2008～2012年度「木簡など出土文字資料積読支援システムの高次化と総合的研究拠点データベースの構築」(課題番号20222002)及び2013～2017年度「木簡など出土文字資料の資源化のための機能的情報集約と知の結集」(課題番号25220401)〈いずれも研究代表者渡辺晃宏〉で一層の進化と完成が目標されている課題でもある。本稿では、上記科研での取り組みを中心に、デジタル化した資料の概要とこれらの活用例を紹介する。

2. デジタル化した資料の概要

過去に集積していたアナログ媒体の資料からデジタル化したものは、(1)記帳ノートと、(2)写真資料と撮影である。

(1) 記帳ノート

当研究所が記帳ノートと称している記録は、木簡1点ずつの形状・加工痕や墨痕を人間の目と手によって観察した記録である。この墨痕の記録から解読結果として、釈文が作られる。木製遺物としての形状の記録と、文字資料としての墨痕の記録という二大要素がこのノートに凝縮されている。実物を収蔵した後々も、報告書作成、あるいは、個別の研究を進めるには、このノートが何度も参照される。

この記帳ノートは、用紙やファイリングの形式は変わりつつも、最初に記録がとられた約50年前から現在に至るまで蓄積されている。その数量もさることながら、長年の使用による紙の劣化も著しくなっていた。そこで、ノートのデジタルスキニングを開始し、現在、過去のものはほぼ完了している。

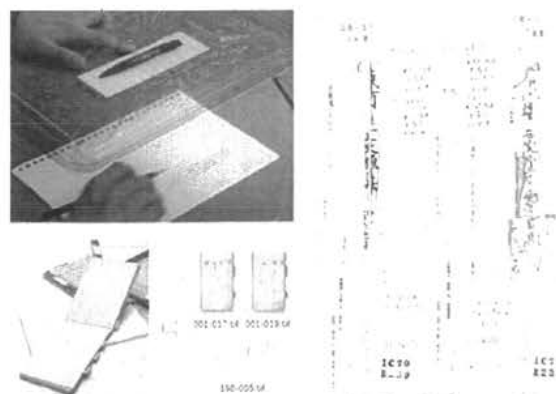


図1 記帳作業と記帳ノート

Figure 1 Recording operation and recording note books.

(2) 写真資料と撮影

①過去の写真資料のデジタル化

^{†1} (独法)国立文化財機構奈良文化財研究所〈アソシエイトフェロー〉
Nara National Research Institute for Cultural Properties (Associate fellow)

a) 発掘状況の詳細は渡辺報告を参照のこと。

木簡の整理・記録作業においては、必ず写真撮影が行われる。撮影は、研究所写真部門の専門カメラマンによって、原寸大で撮影されている。また、一定の整理過程を経ても、再撮影されたり、科学的な保存処理にかけられてから、撮影されたりする場合もある。特に保存処理後に文字がよく見えるようになる木簡もあるので、処理後の撮影データも重要である。

撮影媒体には、60年代の撮影ではガラス乾板が、その後、フィルム撮影が続いた。モノクロ、カラー、赤外線撮影がある。その後、2011年度末に研究所での木簡のデジタル撮影についての指針が示され、デジタルへ完全移行した。過去に撮影されたこれらの媒体自体の保管は永続的に可能とはいえず、デジタル化を進めているところである。ただし、その量は膨大であり、ガラス乾板だけでも約6500点にもなる。ガラス乾板については、ほぼ作業を完了したが、フィルムはまだ完了していない。量と予算の問題に加え、ガラスとフィルムの別、フィルムの中でも判が時期によって異なるため、一律的にデジタル化を進めることは難しい。フィルムの利点もやはりあるが、デジタル化によって得られる利便性やフィルム自体の保存を考えると、デジタル化は急務であるといえる。その一方で、増大するデジタルデータの保存・管理の体制も整えていくべき課題でもある。

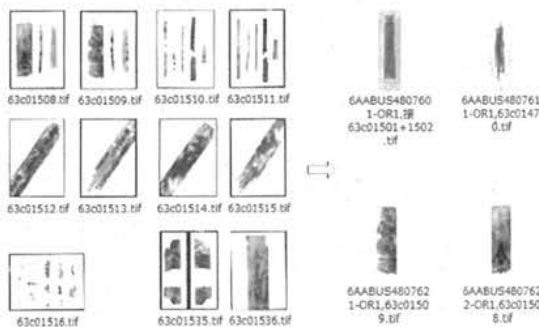


図2 デジタル画像

Figure 2 digital images.

集積例として、図1は、ガラス乾板からデジタル化した画像群である。フィルムの場合でも同様である。この図の左側は、乾板1枚ごとにデジタル化したもので、研究所で付与される写真番号によって管理されている。撮影の過程上、1フィルムに数点の木簡が写し込まれることや、原寸大撮影のために1フィルムに収まりきらないこともある。これらを1点の木簡として利用しやすくするために、図の右側のように、順次、1点ずつ切り分け、あるいは接続して、1木簡を表裏別に個別番号を付与して集積している。

②デジタル撮影と赤外線撮影・観察

木簡のデジタル撮影に適したレンズの開発も進められた。カラー撮影と赤外線撮影の切り替えが容易になり、作業効率が向上した。赤外線による観察は、カメラマンによる上

述の写真撮影とは別に、赤外線モニタを利用した観察が、木簡を担当する研究員によって行われている。表面の腐食や汚損によって、一見文字が見えなくても、赤外線によって解読できる場合がある。この画像データも後々の解読作業のために利用される。

3. デジタルデータの活用

前節で述べたデジタル化したデータを利用して、現在のところ、(1)データベースの作成と公開、(2)字形パターン蓄積と文字認識技術への応用、(3)整理・解読作業への電子システムの導入が試行されている。

(1) データベースの作成と公開

奈良文化財研究所では、上記の画像デジタル化作業に入るよりさらに前に、テキストを中心とする、「木簡データベース[b]」が作成され、現在、公開用と内部業務用を運営している。本文のテキスト検索だけでなく、遺跡名、発掘回数、法量、型式番号等の情報も登録され、いずれも検索できるようになっている。さらに詳細な項目が登録されている内部業務用は、遺物の管理台帳のような役割を担う。

このテキストデータを基礎とし、前掲科研の取り組みから、多くのデータベースが作成、公開されてきた。

- ・「木簡画像データベース 木簡字典[c]」(2005年)
- ・「木簡画像データベース・木簡字典/電子くずし字字典データベース連携検索[d]」(2008年)
- ・「木簡人名データベース[e]」(2011年)
- ・「全国木簡出土遺跡・報告書データベース[f]」(2005年)

このほか、木簡字典のシステムを応用した他の出土文字資料や文書の画像データベースも展開している。

- ・「墨書土器字典[g]」(2011年)
- ・「正倉院文書字典[h]」(開発中)

これらのデータベース群に対する基本的な考え、及び、拠点機能としてのデータベース群の展開と深化については、文献[1][2]を参考にしたい。

また、この他にも、木簡に関する総合情報サイトとして、

- ・「木簡ひろば」[i]

も運営し、展示会情報やコラム、用語集などを掲載、木簡を身近にもらえるよう、情報発信の基盤にしている。

b) <http://www.nabunken.jp/Open/mokkan/mokkan2.html>(1999年)

c) <http://jiten.nabunken.go.jp/index.html>

2014年3月に韓国語、中国語、英語版も公開した。

d) <http://r-jiten.nabunken.go.jp/> 東京大学史料編纂所と連携。

e) http://jinmei.nabunken.go.jp/mokkan_name/ 科学研究費補助金学術創成研究「目録学の構築と古典学の再生-天皇家・公家文庫の実態復原と伝統的知識体系の解明-」(2007~2011年度)および基盤(s)「日本目録学の基盤確立と古典学研究支援ツールの拡充-天皇家・公家文庫を中心に-」(2012~2017年度)(いずれも研究代表者田島公)の成果による。

f) <http://mokuren.nabunken.go.jp/NCPMKR/Mkn-Iseki.html>

g) <http://bokushodoki.nabunken.go.jp/>

h) 奈良女子大学黒田洋子氏の科学研究費補助金基盤研究(C)「書状文化の源流を求めて」(2012~2016年度)との共同研究による成果が含まれている。

i) <http://hiroba.nabunken.go.jp/index.html>

上記のうち、前項のデジタル化した資源を最大限に活用したのが、「木簡字典」であり、前掲3科研における中心的役割を担っている。この「木簡字典」は、上記の「木簡データベース」からより画像に特化し、文字画像の検索ができる(図3)。複数文字検索にも対応し、テキスト中のどの文字列でも取り出せる。これは、各文字を切り出す際に得られる画像タイルデータから合成表示することによって実現している。その一方で1文字画像も自動的に集積している。

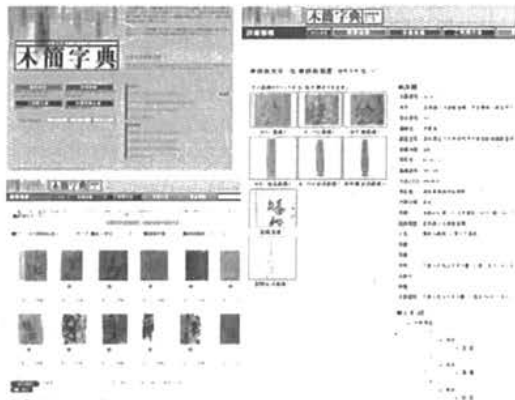


図3 画像データベース 木簡字典
Figure 3 The Wooden Tablet Image Database.

同一木簡の画像は、10点まで表示でき、モノクロ画像、カラー画像、赤外線画像や、前述の記帳ノート画像を同時に閲覧できる。これは、多くの画像を利用者に提供するという目的だけではなく、木簡の解読には、あらゆる写真材料から検証するというノウハウを反映した機能でもある。

また、「木簡字典」は、東京大学史料編纂所「電子くずし字字典データベース」との機能をこえた連携によって、飛鳥・奈良時代から江戸時代に至るまでの1000年を越える、文字の変遷を手軽に見ることができるようになった。

そして、「木簡字典」で表示される情報の中には、「木簡データベース」の基本情報のほかに、様々な機能が順次搭載されている。特に、以下の2点の機能のコンセプトと実現は、馬場基科研の成果によるものである[j]。1点は、基本情報欄から、関連する情報へのリンクを貼る編集機能である。これにより、利用者は、基本情報から、各自の興味にあわせて、WEB上の他の情報を得ることができる。例えば、古代の地名から現在該当する地点の地図を表示するなどである。もう1点は、基本情報の他に、木簡の本文テキストに付けられたタグを明示し、各語句がどのような意味、部類の語句か、利用者に分かりやすく表示できる機能で、テキストデータにタグ付けしたXMLデータを集積してきたことによって実現されている。この表示部分の各タグ名

にもリンクを貼ることができ、そのタグに関連する情報へもたぶることができるようになっている。木簡を対象としたタグの選別からタグ付けの作業には、これまでの木簡研究の成果が盛り込まれており、それがデジタル上で表現された意義は大きい。XMLデータに取り込まれた画像上の位置情報を応用した語句の出現位置の統計など、デジタルデータと技術を応用した研究がすでに展開されている。

ところで、以上のリンク機能で紹介したリンク先の多くは、木簡ひろばに集積された用語解説やコラムになっている。また、木簡ひろばの管理システム内で、語句を登録し、リンク先を指定すると、「木簡字典」タグ表示の該当語句に、自動的にリンクが貼れる。「木簡ひろば」は、これまでの渡辺科研で、知識データベースの拠点となってきた。この他にも、木簡研究文献データベースを作成中であり、木簡に関する知識や関連文献へ容易に辿りつけることができるようになっている。また、人名データベースの遺構の年代観表示は、報告書等からの考古学的知見・記述のデータを集積した結果であり、これ自体がデータベースとなっている。

(2) 字形パターンの蓄積と文字認識技術への応用

木簡の表面は、腐食が進んでいたり、木材が折れていたり、文字が半裁されていたりして、不十分な墨痕が多く、解読に困難をもたらす。このような場合には、蓄積してきた類例や知識が駆使され、解読に至る。墨痕についていえば、以前にみた筆画のパターンがヒントとなる。あるいは、類例がまだあまりなかった、木簡が発見された初期の頃の木簡も、類例が増えたことによって解読される場合もある。これまで担当者が脳裏に蓄えてきた類例、および、蓄えていく作業を、デジタル上で、いつでも検索し、画像を具体的に表示するというのが「木簡字典」でもある。

そして、この文字認識をはじめとする、木簡解読に必要な知識や認識処理から解読結果活字で表現するまでの過程を、システム上で再現しようと開発されたのが、木簡釈読支援システム「Mokkanshop[k]」である。Mokkanshopは、東京農工大学中川研究室によって、奈良文化財研究所での解読支援のために開発されたツールで、木簡ひろば釈読室のページから提供している。開発の経緯や基本的なシステムの紹介については文献[3]を、現在に至るまで、各技術の研究が重ねられた成果は、文献[4]を参照いただきたい。機能を簡単にまとめると、次の4点である。

- ・画像処理機能;腐食や木目で見えにくい画像から墨痕の抽出に有効な画像処理をより簡単にする。また、折損によって筆画が失われた部分を補う。
- ・文字認識機能;処理した画像から文字を認識し、搭載した文字画像(辞書)の中から、候補を挙げる。また、候補からWeb上の木簡字典も参照できる。
- ・文脈処理機能;関連する文脈の候補を挙げる。

j) 科学研究費補助金若手研究(B)「木簡の構文・文字表記パターンの解析・抽出研究」(2008～2010年度)。および、「GT-Map 等時空間解析システムを利用した木簡等出土文字資料分析の基礎的研究」(2008～2010年度)。

k) 2010年3月5日に商標登録第5307067号が取得されている。

- ・積文パネル機能:文字候補から選び出して解読を進めた結果を活字として記録,保存する。記録データは,他の利用者とも共有できる。

特に,文字認識機能には,これまでの文字画像データから類似するパターンを検出するものであり,木簡解読の担当者の脳裏を再現するものである。この脳裏を支える辞書を一層強化していくことによって,さらに解読の際の候補の幅が増えていくことが期待される。



図 4 木簡積読支援システム MokkaShop の GUI

Figure 4 GUI of MokkaShop

MokkaShop は,木簡字典データベースとともに,前述の科研における二大拠点機能の1つであり,現在も,末代誠仁氏を中心として,認識率の向上のための技術開発と機器の進化に適合するインターフェイスの改良が行われている。

(3) 整理・解読作業への電子システムの導入

2.(1)で述べた写真データの集積については,整理・解読作業の過程で記録と解読の資源となるものであった。しかし,2008年に,膨大な出土した削屑に対し,上述のような1点1点の記録や撮影が困難になった状況を受け,整理作業,遺物の管理の効率化が目指された。そこで東京農工大学中川研究室によって開発されたのが,木簡アノテーションツールである。本システムの開発の経緯と今後の展開については,馬場報告[1]を,ツールの機能詳細については,文献[5][6]を,参照していただきたい。

ここで簡単に紹介すると,これは,デジタルカメラで撮影した画像を出発点とし,アノテーション技術によって,必要な情報を付与できるツールである。このツールでは,下図のように,ガラス板にのせられた削屑10点を1単位とする画像上で,整理作業が進行する。特に,各10点の範囲を識別し,R番号という整理上必須の番号を自動的に割り振ってくれるという大変便利なツールである。主な機能は次の通りである。

- ・画像及びメタデータ表示
- ・木簡領域描画及び編集
- ・注釈
- ・自動R番号配置
- ・画像処理
- ・メタデータ出力
- ・検索

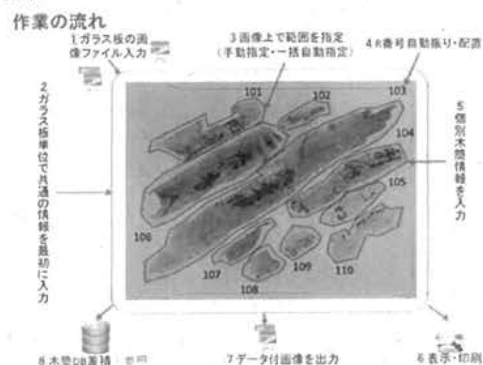


図 5 木簡アノテーションシステム MokkaAnoter

Figure 5 Mokkan annotation system MokkaAnoter.

4. おわりに

以上,木簡の整理・解読作業におけるデジタルデータの利活用と電子システムの導入について紹介した。引き続き,木簡の解読や整理の実態に即した技術の応用が望まれる。

謝辞

本稿は,科学研究費補助金基盤研究(S)「木簡など出土文字資料の資源化のための機能的情報集約と知の結集」(2013-2017年度,研究代表者渡辺晃宏,課題番号25220401)の成果による。

参考文献

- 1) 渡辺晃宏,馬場基,井上幸:出土文字資料の画像データベースの構築,奈良文化財研究所紀要2012, pp.54-55(2012).
- 2) 馬場基,渡辺晃宏,井上幸,中川正樹,末代誠仁:個別DBの深化と連携確保をめぐる,研究報告人文科学とコンピュータ(CH),2013-CH-97(10), pp.1-4(2013).
- 3) 末代誠仁,中川正樹,馬場基,渡辺晃宏:古代木簡解読支援のための画像処理および字体検索の高度化,じんもんこん2011論文集2011(8), pp.93-98(2011).
- 4) 末代誠仁,白井啓一郎,遠藤友樹,中川正樹,馬場基,渡辺晃宏,井上聡,久留島典子:古代木簡に対する平滑化処理の適用および古代木簡解読支援システムのアップデート,じんもんこん2013論文集2013(4), pp.65-70(2013).
- 5) PhanVanTruyen,中川正樹,馬場基,渡辺晃宏:木簡整理のためのアノテーションツールの開発,じんもんこん2011論文集2011(8), pp.249-254(2011).
- 6) PhanVanTruyen,中川正樹,馬場基:木簡画像集録システムの設計と実現,情報考古学(日本情報考古学会誌)19(1・2), pp.1-12(2014).

1) 馬場報告2.整理・調査ツールとしての期待の項参照のこと。