

科学研究費助成事業（基盤研究（S））公表用資料
〔平成28年度研究進捗評価用〕

平成25年度採択分
平成28年3月30日現在

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

Central information management and knowledge compilation
for the development as a resource of wooden tablets and
other excavated documentary materials

課題番号：25220401

渡辺 晃宏 (WATANABE AKIHIRO)

独立行政法人国立文化財機構奈良文化財研究所・都城発掘調査部・副部長



研究の概要

木簡など出土文字資料のもつ情報抽出の効率化を図るとともに、これを一元的に管理して資源化し、結集した知を新たな資料の情報抽出に役立てる知の循環を確立する。そして、出土文字資料統合データベースの構築を図り、海外を含めた他機関と連携しつつ、日本の木簡の約7割を調査・保管する機関として相応しいセンター機能の構築・遂行に資する。

研究分野：人文学、史学、日本史、史料研究

キーワード：日本史、史料研究、木簡学、出土文字資料、漢字、文字認識

1. 研究開始当初の背景

奈良文化財研究所では総合的・効率的に木簡情報を提供する研究拠点ツールとして、釈読支援システム「Mokkanshop」と木簡画像DB「木簡字典」を作成・公開してきた。

その過程で、木簡に関する研究の拠点的機能をさらに強化するために、a 膨大な木簡資料の情報取得の効率化、b 木簡資料に関するさまざまな知の結集、c そうして集められた木簡に関する情報や知を効率的に保管し活用するためのシステムの確立、以上3つの課題が明らかになってきた。

2. 研究の目的

上記の課題を解決することで、歴史叙述に重要な役割を果たすようになった木簡など出土文字資料のもつ情報を一元的に管理して資源化し、結集した知を新たな資料の情報抽出に役立てる知の循環を確立したい。

また、日本の木簡の約7割を調査・保管する機関に見合うセンター機能に資するとともに、木簡など出土文字資料が、国民共有の身近な生きた歴史資料となるよう努めたい。

3. 研究の方法

上記の3つの課題に対応し、A 木簡など出土文字資料のもつ情報を効率的に抽出・管理する研究、及びそれらの画像データの拡充、B 木簡そのものやその周辺の関連情報を効率的に活用できるような知を結集する研究を基礎に、C 釈読支援システム「Mokkanshop」と木簡画像データベース「木簡字典」を中核に据えた統合木簡データベースを構築する

4. これまでの成果

A、木簡など出土文字資料情報抽出の効率化
(1)資料の情報を効率的に管理・活用するシステムの開発

・アノテーションツールの開発・運用 木簡の情報を一元的に管理するためのアノテーションソフト (Mokkanotator) のB版を実用化した。削屑整理の大幅な効率化・省力化が図れ、『平城宮発掘調査出土木簡概報(43)』(2015/11刊)の編集に生かすことができた。

(2)情報そのものの拡充

・既撮影フィルムのデジタル化と新規撮影 木簡の既撮影資料(フィルム)のデジタル化と、平城宮・京跡出土木簡の新規カラー撮影を進め、計約5,500点(表裏別カウント)の木簡の約3万文字画像を「木簡字典」にアップした(累計約14,000点、9万文字画像)。
・全国出土木簡等の撮影 長門鑄銭所跡、伊場遺跡群、秋田・鳥取・埼玉県内などの木簡・墨書土器の撮影を継続して実施した。

B、木簡など出土文字資料の知の結集

(1)画像検索機能の構築

・木簡・くずし字解読システム—MOJIZO—の開発・公開 従来の「木簡字典」は、テキスト入力(文字引き)による文字画像データベースだったが、今回新たに画像入力(画像引き)による文字画像データベースを構築・公開した。これは先行科研で開発した釈読支援システム「Mokkanshop」(モッカショップ)を高次化し、Web用プログラムに改変したものである。同時に東京大学史料編纂所の電子くずし字字典DBとの連携も図り、テキストによる検索に対して画像による検索シ

システムとして生まれ変わらせることに成功した。今後未解読資料や、記号など非文字資料の検索への応用が期待できる。

(2)木簡の文字の字体研究の深化

・木簡所用文字の標準字体の検討 標準字体一覧を作成すべく、異体字関係を検討中しつつデータ抽出作業を進めている。

(3)検索機能の強化

・「木簡字典」の高次化 語句を意味分類シタグ付けした XML データを利用し、意味ツリーから選択・検索する意味検索機能を搭載した結果、異表記の検索が可能になった。

(4)周辺データベース群の充実

・木簡研究論文DB 個々の木簡に関する研究文献を『木簡研究』掲載論文などから抽出して「木簡研究文献一覧」として整理し、「木簡字典」に搭載して公開した。また、「日本古代史研究文献目録データベース」(法政大学国際日本学研究所)と国立情報学研究所「CiNii」へのリンクを貼った

・「古代地名検索システム」 Mookanshop に搭載していた文脈処理機能のうち、地名検索について、漢字と読みから検索できる単独システムを開発・公開した。

・「正倉院文書字典」 奈良女子大学古代学学術研究センターの黒田洋子氏の協力を得て、「木簡字典」の枠組みを応用した「正倉院文書字典」のクライアントPC版を開発した。未公開だが、Web 公開システムも構築した。C、出土文字資料統合データベースの構築と連携の拡充

(1)統合データベースの構築 「木簡・くずし字解読システム—MOJIZO—」により、テキスト検索と画像検索を兼ね備えた統合木簡データベース実現への見通しを得た。

(2)情報発信や他機関との連携

木簡の実物展示「地下の正倉院展」を毎年秋に継続して開催した。また、台湾中央研究院歴史語言研究所との間で、木簡・簡牘の研究協力に関する協約書を締結し、同所所蔵の居延漢簡の赤外線撮影を行った。

以上、これまでの3年間の研究で、木簡など出土文字資料に関する知の結集と資源化の枠組み完成に向けた歩みを軌道に乗せることができたと考える。

5. 今後の計画

A 木簡など出土文字資料の情報抽出の効率化について、アノテーションツールの削屑以外の木簡への応用、B 木簡など出土文字資料の知の結集について、①削屑の形状による検索機能の構築、②木簡の標準字体一覧の作成と文字認識精度向上への活用、③「正倉院文書字典」の Web 公開、④古代地名検索システムへの木簡表記の搭載などを進める。

以上を受け、C 出土文字資料統合データベースの構築と連携の拡充について、従来の「文字引き」と「木簡・くずし字解読システ

ム—MOJIZO—」の「画像引き」を統合した統合木簡データベースを構築し、β版の公開をめざす。また、引き続き諸外国・機関との連携を図り、東アジア木簡学・資料学に寄与したい。当初の研究目標は概ね達成できる見込みだが、今後広い情報発信や学校教育への活用などにも配慮していきたい。

6. これまでの発表論文等(受賞等も含む)

1、渡辺晃宏「日本歴史学における木簡研究成果」日韓木簡ワークショップ発表論文集『木簡研究の成果』(ソウル大東洋学研究所教授李丞宰編、p.p.71-98〈韓文〉、p.p.197-219〈日文〉)、2016年

2、Combination of Global and Local Contexts for Text/Non-text Classification in Heterogeneous Online Handwritten Documents、Truyen Van Phan and Masaki Nakagawa、Pattern Recognition51、p.p.112-124、2016

3、未代誠仁、馬場基、渡辺晃宏、井上聡、久留島典子、中川正樹「古文書字形デジタルアーカイブのための検索システムの試作」『じんもんこん 2015 論文集』p.p.9-15、2015年

4、未代誠仁、白井啓一郎、馬場基、渡辺晃宏、井上聡、久留島典子、中川正樹「古文書デジタルアーカイブに対する横断的字形検索サービスの試作」『じんもんこん 2014 論文集』p.p.87-92、2014年

5、渡辺晃宏「墨書のある木製品とその機能—東アジア木簡学の確立のために—」角谷常子編『東アジア木簡学のために』汲古書院、p.p.251-271、2014年

6、Keiichiro Shirai、Yuuki Endo、Akihiro Kitadai、Satoshi Inoue、Noriko Kurushima、Hgajime Baba、Akihiro Watanabe、Masaki Nakagawa、Character Shape Restoration of Binarized Historical Documents by Smoothing via Geodesic Morphology、Proc. 12th International Conference on Documents Analysis and Recognition1、p.p.1317-1321、2013

7、特許出願。出土文字資料解読システム及び装置、馬場基、渡辺晃宏、方国花、高田祐一、未代誠仁、特願 2016-164号。

8、特許出願。文字認識方法及び文字認識装置、渡辺晃宏、馬場基、高田祐一、未代誠仁、特願 2015-243661号。

9、特許出願。出土文字資料解読システム及び装置、渡辺晃宏、馬場基、高田祐一、井上幸、未代誠仁、特願 2015-234684号。

(7~9の権利者は、いずれも独立行政法人国立文化財機構)

ホームページ等

<http://mojizo.nabunken.go.jp/> 木簡・くずし字解読システム—MOJIZO—

<http://chimei.nabunken.go.jp/> 古代地名検索システム