

平成17年度科学研究費補助金（基盤研究（S））研究状況報告書

ふりがな（ローマ字）		WATANABE AKIHIRO					
①研究代表者 氏名		渡辺 晃宏		②所属研究機関・ 部局・職			
				(独) 奈良文化財研究所・平城宮跡 発掘調査部・史料調査室長			
③研究 課題 名	和文	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					
	英文	The development of the automatical system with a reasoning function which recognizes the character of the excavated written sources such as Wooden Tablets					
④研究経費		平成15年度	平成16年度	平成17年度	平成18年度	平成19年度	総合計
17年度以降は内約額 金額単位：千円		11,900	17,300	15,900	13,300	16,200	74,600
⑤研究組織（研究代表者及び研究分担者）							
氏名		所属研究機関・部局・職		現在の専門		役割分担（研究実施計画に対する分担事項）	
渡辺 晃宏		(独) 奈良文化財研究所・平城宮跡 発掘調査部・史料調査室長		日本古代史		研究代表者として研究全体の統括	
金子 裕之		(独) 奈良文化財研究所・飛鳥藤原 宮跡発掘調査部・部長		考古学・古代都城制		釈読支援データベース群の研究（考古部門）	
綾村 宏		(独) 奈良文化財研究所・文化遺産 研究部・部長		日本中世史		釈読支援データベース群の研究（歴史部門）	
森本 晋		(独) 奈良文化財研究所・埋蔵文化 財センター・主任研究官		日本考古学・情報学		文字画像データベースの研究	
馬場 基		(独) 奈良文化財研究所・平城宮跡 発掘調査部史料調査室・研究員		日本古代史		文字画像データベースの研究・木簡データ取り込みシステムの研究	
山本 崇		(独) 奈良文化財研究所・平城宮跡 発掘調査部史料調査室・研究員		日本古代史		文字画像データベースの研究・釈読支援データベース群の研究（歴史部門）	
吉川 聡		(独) 奈良文化財研究所・文化遺産 研究部歴史研究室・研究員		日本古代史		釈読支援データベース群の研究（歴史部門）	
大河内 隆之		(独) 奈良文化財研究所・埋蔵文化 財センター古環境研究室・研究員		年輪年代学・情報学		木簡画像鮮明化システムの研究	
山田 奨治		国際日本文化研究センター・教授		情報学		釈読支援データベース群の研究	
柴山 守		京都大学・東南アジア研究所・教授		情報工学		木簡用OCRの開発研究	
及川 昭文		総合研究大学院大学・教育研究情報 資料センター・教授		数理考古学		釈読支援データベース群の研究	
中川 正樹		東京農工大学・工学部・教授		情報科学		木簡用OCRの開発研究	
小口 雅史		法政大学・文学部・教授		日本古代史		釈読支援データベース群の研究	
鈴木 卓治		国立歴史民俗博物館・情報資料研究 部・助手		博物館情報システム		木簡データ取り込みシステムの研究	
⑥当初の研究目的（交付申請書に記載した研究目的を簡潔に記入してください。）							
<p>1961年に平城宮跡で初めて木簡が出土してから42年、この間に全国で見つかった木簡は20万点を超えた。奈良文化財研究所では、このうち約8割の16万点強の木簡を発掘し、調査・研究を続けてきた。木簡研究の基礎となる最も重要な作業は文字の読み取り、釈文の作成であるが、これは熟練した研究員の長年にわたる経験と知識、そして勘に頼らざるを得ない部分の多い作業である。本研究では、こうした長年の釈読技術の蓄積を生かしつつ、近年めざましい進展を遂げつつある古文書における手書き文字OCR技術による文字自動認識システムを応用して、木簡など出土文字資料の文字の読み取り、釈文作成作業を自動的に行う方法を開発し、もって釈読技術の汎用化と効率化を図り、ひいては奈良文化財研究所のもつ膨大な文字資料を画像とともに検索に耐え得るデータとして蓄積し、広く研究者や一般の便に供することを目的とする。</p> <p>開発を目指すシステムは、概ね以下の通りである。</p> <ol style="list-style-type: none"> ①木簡情報デジタル化簡易システム—木簡の文字情報を、赤外線撮影の利用や画像処理による木目など余剰な情報との区別などにより、できるだけ鮮明にデジタル化して蓄積するシステムを開発する。 ②木簡の「文字」の事典—個々の文字について、サンプルとなる画像をセットにしたデータベースを作成する。 ③木簡釈読支援システム—文字データベースとともに横断的に検索することによって木簡の釈読を支援するデータベース群を構築する。 ④木簡用OCR—文字の劣化や欠損により不完全な文字を類推して釈読するシステムを開発する。 <p>①～④により、客観的かつ汎用性のある、推論機能をもつ木簡釈読システムの提示を目指したい。</p>							

⑦これまでの研究経過（研究の進捗状況について、必要に応じて図表等を用いながら、具体的に記入してください。）

研究目的記載の①～④の4項について、研究の進捗の著しいものから順に述べる。

a 文字画像データベースの開発（研究目的②に相当）

ア、文字画像切り出しソフトの開発

木簡写真から文字を一文字ごとに切り出し、蓄積するためのソフトを開発した。そして、既存の4×5カラーリヴァーサル写真・モノクロ写真、赤外線写真、及び後述の記帳ノートのデータについての文字画像の切り出しを行い、2年間で延べ約6,300文字分の画像を蓄積した。このデータは後述の木簡画像データベース「木簡字典」に利用した。

イ、木簡記帳ノートの電子化及び文字の切り出し

木簡を解読した際に作成した見取図及びデータを記入したノート（記帳ノート）の電子化を行い、これを切り出し対象とすることとした。奈良文化財研究所が調査した木簡のうち、平城宮跡発掘調査部担当分の約三分の二（約10,000頁分）についてカラーマイクロ撮影を行い、木簡ごとの切り出しを実施した。成果はフィルム・頁ごとのデータ・木簡ごとのデータ・記帳ノートの複本の形で得ている。

ウ、「木簡字典」のWEB公開（2005年2月8日奈良文化財研究所のホームページ上で）

「木簡字典」は、木簡に書かれた文字ごとの画像データベースで、文字種ごとに実際に書かれた字体の事例を、モノクロ写真だけでなく、カラー・赤外線、さらには私たちが解読した記録（記帳ノート）も含めた複数の画像で紹介する画期的なシステムである。また、従来から奈良文化財研究所が公開している木簡データベースのデータを用いて、その画像の文字がどのような文脈で用いられたかを含め、その文字が書かれた木簡そのものの基礎データが全てわかるようになっている。公開した文字画像はアで切り出した延べ約6,300文字分に及ぶ。

木簡字典はNHKニュースや朝日新聞、読売新聞でも取り上げられ、公開から約50日でアクセス件数は、約2,500件に及び、画期的なシステムとして好評を博している。

b 文字自動認識システム（OCR）の開発（研究目的④に相当）

OCR用に必要な情報のあり方を検討して切り出しソフトに反映させ、基礎データの蓄積に努めた。また、既存の文字自動認識システムの現状についての研究者間の理解を深め、木簡など出土文字資料への利用についても一定の見通しを描くことができた。

具体的には木簡の文字解読支援システム「mokkan shop」を試験開発した。

「mokkan shop」はオフラインの文字認識処理システムで、文字画像の切り出し、墨部の抽出、文字認識、認識結果の検証という手順で、可能性の高いものから順に認識候補を表示して木簡解読を支援するものである。これまでに認識対象とできた文字パターンは約241字種1,270パターンである。墨の部分の抽出するための画像処理手法、及び欠損文字の認識についても有効な文字認識システムの開発に着手し、その有効性を確認することができた。

c 木簡解読のための支援データベースの構築（研究目的③に相当）

古代地名データベース、古代人名データベースなど、構築すべきデータベースの選定、ならびに入力すべき具体的な資料の検討を実施し、下記の2項目について入力を行った。

ア、地名データベース『和名類聚抄』に見える国・郡・郷名（郷レベルで約4,040件）のデータ入力を完了した。

イ、古代人名データベース日本古代の人名約26,600名について、人名と居住地域名のデータ入力を行った。

d 文字画像鮮明化のためのシステムの開発（研究目的①に相当）

ア、木目などのノイズ除去ソフトの開発

β版ソフトの開発を終了し、引き続き改良を実施している。また、これを実際に運用するためのデータ収集を、特定の木簡を素材として行った。今後これをさらに汎用化し完成を目指す。

イ、デジタルカメラなどの有効性の実験

簡易な赤外線画像が得られる機器として注目されるデジタルカメラについて、実際の木簡の調査において有効性を確認した。リアルタイムでの画像を得るという点では既往の赤外線デジタルカメラ装置には劣るが、赤外線画像を得る効果についてはその利用に一定の目途を得た。また、デジタルビデオカメラの利用についても、一定の目途を得た。デジタルカメラの静止画像の利用とともに、その有用性をさらに高めていくことが可能と思われる。

⑧特記事項 (これまでの研究において得られた、独創性・新規性を格段に発展させる結果あるいは可能性、新たな知見、学問的・学術的なインパクト等特記すべき事項があれば記入してください。)

本研究は、木簡など出土文字資料の文字に関する自動認識システムの開発をメインに据えながら、これを核にもつ木簡などの文字を解読するための総合的な積読支援システムの構築を目指すものである。従来の歴史資料のOCR研究とは根本的に異なる次のような特徴を備えている。

- 1、出土文字資料という必ずしも完存しない文字のデータを対象としていること。墨痕の消失や欠損によって字画の一部を欠いている可能性のある文字を対象としていること。
- 2、機械に文字を読みとらせるだけでなく、研究者との対話を行いながら文字積読を進めていくシステムであること。
- 3、OCRのシステムとして独立したものではなく、他のさまざまな木簡積読支援システムと有機的に結合した総合的な出土文字資料積読支援システムの構築をめざしていること。
- 4、これまでに研究者が長年にわたって蓄積してきた出土文字資料積読の経験が生かされたシステムであること。
- 5、一部の研究者のためのシステムではなく、広く出土文字資料の解読に志す人々に資するツールたり得るシステムをめざしていること。

2004年度に奈良文化財研究所のホームページ上で公開することができた木簡画像データベース「木簡字典」は、上記の木簡積読支援システムの一翼を担うものであり、将来的には既存の木簡データベースを包摂する総合的な木簡データベースとして機能することが期待されるものである。その特徴は次の通りである。

- 1、従来の木簡1点ごとのテキストデータを基調とするデータベースとは異なり、木簡の文字一文字ごとの画像のデータベースであること。
- 2、モノクロ写真だけでなく、カラー写真、赤外線デジタル写真など、異なる撮影方法による画像が複数見られること。
- 3、写真だけでなく、奈良文化財研究所の研究員が木簡を解読した記録である記帳ノートも公開していること。
- 4、当該文字画像を含む木簡ごとのテキストデータを参照させることによって当該文字画像が木簡のどのような文脈で用いられたものであるかが、瞬時にしてわかるようになっていること。
- 5、研究者だけでなく、一般のユーザの利用をも想定した「簡単検索」を設けるなど、データベースとしてのユーザ・インタフェースに配慮したものとなっていること。特に、「簡単検索」のうちの、木簡の形状の絵からの検索や、地図による国名検索は、歴史資料データベースの検索のあり方を見直す画期的なシステムである。
- 6、従来の検索システムは不可能であった木簡の大きさの範囲指定が可能になるなど、検索の便が大幅に向上したこと。

これまで中国の法書や碑文などを素材とした漢字のさまざまな字体を例示する字典はあったが、日本の資料、それも木簡のような生の史料によって漢字の字体を例示する字典、しかもデータベースとしての公開は私たちの「木簡字典」が初めてである。また、当該文字を前後の文脈の中で理解できるような配慮は従来全くなされてこなかったが、木簡のような生の史料の場合、前後の文脈のもつ意味は大きく、これを捨象することはできない。「木簡字典」では、単なる文脈だけでなく、当該文字を含む木簡そのもののデータを随時参照できるようになっており、この点は画期的なことである。これは歴史資料データベースとしても全く新しい形を提示したものである。

⑨研究成果の発表状況 (この研究費による成果の発表に限り、学術誌等に発表した論文(掲載が確定しているものを含む。)の全著者名、論文名、学協会誌名、巻(号)、最初と最後のページ、発表年(西暦)、及び国際会議、学会等における発表状況について記入してください。なお、代表的な論文3件に○を、また研究代表者に下線を付してください。)

【論文など】

- 奈良文化財研究所(平城宮跡発掘調査部編、担当渡辺晃宏・古尾谷知浩)、『平城京漆紙文書1』、2005年、本文80頁、カラー図版2葉、図版28葉((財)東京大学出版会からの市販分もあり)
- 齋藤恵・蜂谷大翼・未代誠仁・中川正樹・馬場基・渡辺晃宏、「木簡画像から墨の部分の抽出するための画像処理手法」、『電子情報通信学会技術報告』PRMU2005-03-18、2005年、p.p.163-168
- 蜂谷大翼・齋藤恵・未代誠仁・中川正樹・馬場基・渡辺晃宏、「欠損を含む文字パターンを対象とした文字認識システムの試作」、『電子情報通信学会技術報告』PRMU2005-03-19、2005年、p.p.169-174
- 渡辺晃宏、「駿河国正税帳の世界一天平年間の財政文書と古代の藤枝」、『藤州市史研究』6、p.p.49-72、2005年
- 渡辺晃宏、「書評平川南著『古代地方木簡の研究』」、古文書学会編『古文書研究』60、頁未定、2005年
- 山本崇・高橋学、「鷹巣町胡桃館遺跡出土の木簡」、『秋田県埋蔵文化財センター研究紀要』19、2005年、p.p.69-82
- 榎本剛治・山本崇、「特集『胡桃館遺跡』37年目の大発見!」、『広報たかのす』2005年3月1日号、2005年、p.p.8-11
- 山本崇、「秋田県・胡桃館遺跡出土木簡の釈読」、奈良文化財研究所『奈文研ニュース』NO.16、2005年、p.7
- 奈良文化財研究所平城宮跡発掘調査部史料調査室(担当山本崇)『大極殿関係史料(稿)(二)編年史料』、2005年、本文376頁
- 奈良文化財研究所(平城宮跡発掘調査部編、担当渡辺晃宏)、『平城宮木簡6』、2004年、本文580頁・図版112葉(明新印刷からの市販分もあり)
- 奈良文化財研究所埋蔵文化財センター(山本崇・寺崎保広共編・責任編集)、『全国木簡出土遺跡・報告書総覧(埋蔵文化財ニュース114)』、2004年、280頁
- 未代誠仁・齋藤恵・蜂谷大翼・中川正樹・馬場基・渡辺晃宏、「木簡解読支援システムの基本設計と試作」、『人文科学とコンピュータシンポジウム論文集』5-A-1、2004年、p.p.215-220
- 奈良文化財研究所(平城宮跡発掘調査部史料調査室編、担当馬場基)、『平城宮発掘調査出土木簡概報37』、2003年、本文37頁・図版12頁
- 渡辺晃宏、「籍帳制」、平川南・沖森卓也・栄原永遠男・山中章編『文字と古代日本1支配と文字』吉川弘文館刊所収、2004年、p.p.116-146
- 渡辺晃宏、「奈良・平城宮跡」、木簡学会『木簡研究』26、2004年、p.p.238-241
- 渡辺晃宏、「奈良文化財研究所における木簡記帳ノートのカラーマイクロ撮影」、『月刊IM』2004年8月号、2004年、p.p.1-7
- 渡辺晃宏、「東区朝堂院南辺官衙の変遷と出土木簡」奈良文化財研究所『平城宮木簡6』、p.p.43-56、2004年
- 渡辺晃宏、「式部省関係考選木簡の特徴」奈良文化財研究所『平城宮木簡6』、p.p.56-62、2004年
- 渡辺晃宏、「木簡にみる古代日本の文字—研究素材の提供者の立場から」奈良女子大学・国際高等研究所共同研究公開シンポジウム「古代日本語をよむ—東アジアの文字環境」第2回、p.p.29-52、2004年
- 馬場基、「宮城の警護」、平川南・沖森卓也・栄原永遠男・山中章編『文字と古代日本2文字による交流』吉川弘文館刊所収、2004年、p.p.272-295
- 馬場基、「木片の調達」、木簡学会『木簡研究』26、2004年、p.231
- 馬場基、「奈良・旧大乘院庭園」、木簡学会『木簡研究』26、2004年、p.13
- 馬場基、「陸奥国荷札の発見」、『奈良文化財研究所紀要2004』、2004年、p.29
- 馬場基、「一行書きの隠岐国荷札」、西洋子さん還暦記念論集刊行会編『洋洋福寿—正倉院文書の部屋—』所収、2004年、p.p.201-207
- 山本崇、「御齋会とその舗設—大極殿院仏事考」、『奈良文化財研究所紀要2004』、2004年、p.p.34-37

- 山本崇、「2003年出土の木簡」、木簡学会『木簡研究』26、2004年、p.p.1-6
- 山本崇、「奈良・法華寺」、木簡学会『木簡研究』26、2004年、p.12
- 金子裕之、「都城における山陵—藤原・平城京と喪葬制」、考古学研究会『文化の多様性と比較考古学』、2004年、p.p.141-150.
- 金子裕之、「宮室・京の成立・大極殿・朝堂院・三部世界観」、奈良文化財研究所『古代の官衙遺跡Ⅱ 遺物・遺跡編』、2004年、p.p.116-123
- 金子裕之、「古代都城の園林に関する基礎的研究—平城京の寺院園林」、韓国忠南大学校百済研究所『古代都市と王権』（韓文）、2004年、p.p.237-264
- 鈴木卓治、「錦絵資料画像の測色値画像データベースの構築」、『日本色彩学会誌』Vol.28,Supplement、2004年、p.p.152-155
- 鈴木卓治・安達丈夫・大久保純一・小林光夫、「錦絵画像の測色値画像データベースの構築と色彩分析の試み」、『人文科学とコンピュータシンポジウム論文集』Vol.2004、2004年、p.p.75-82
- KOBAYASI,Mituo, SUZUKI,Takuzi,YAMAGUCHI, Takeshi、「Conservation of Color Information of Historical Materials Using Standard Digital Cameras」、『Proceeding of the Color and Cultural Heritage 2004』、2004、p8
- 小口雅史、「日本史史料とパソコン—電子史料学の提唱・再論」、『中世総合資料学の可能性—新しい学問体系の構築に向けて』、2004年、p.p.186-200
- 渡辺晃宏、「平城宮第一次大極殿の成立」、『奈良文化財研究所紀要 2003』、2003年。p.p.18-19
- 渡辺晃宏、「公廨の成立—その財源と機能」、笹山晴生編『続日本律令制論集』上、吉川弘文館刊所収、2003年、p.p.143-167
- 馬場基・渡辺晃宏、「奈良・平城宮跡」、木簡学会『木簡研究』25、2003年、p.p.7-14
- 山本崇、「奈良・興福寺一乗院跡」、木簡学会『木簡研究』25、2003年、p.p.17-18
- 山田奨治・柴山守、「n-gram と OCR による定型表現がある古文書の文字の推定」、『情報処理学会研究報告』、2003、59、2003年、p.p.17-24
- 小口雅史・家辺勝文・鈴木卓治、「日本古代資料集の高精細全文テキストデータベースと検索システムの開発—青森県史資料編古代1・同補遺全文データ CD-ROM を事例として」、『情報処理学会シンポジウムシリーズ（じんもんこん 2003 論文集）』Vol.2003, No.13、2003年、p.p.235-242

【学会報告など】

- 渡辺晃宏、「平城宮木簡のデジタルアーカイブ」、平成16年度デジタルアーカイブ構想事業『地域産業関連技術等の高精細デジタル映像によるアーカイブソフト』成果発表会in奈良・基調講演、於奈良県文化会館、2005年3月8日
- 未代誠仁・齋藤恵・蜂谷大翼・中川正樹・馬場基・渡辺晃宏、「木簡解読支援システムの基本設計と試作」、人文科学とコンピュータシンポジウム「じんもんこん 2004」『デジタルアーカイブ—デジタル学術情報資源の共有と活用—』セッション 5A：デジタル考古学、於立命館大学、2004年12月10日
- 馬場基、「「都市」平城京の可能性と限界」、シンポジウム「東アジア古代都市論—古代都城の社会＝空間構造を探る」、於東京大学、2004年11月20日
- 渡辺晃宏、「木簡にみる古代日本の文字—研究素材の提供者の立場から」、奈良女子大学・国際高等研究所共同研究公開シンポジウム「古代日本語をよむ—東アジアの文字環境」第2回、於奈良女子大学、2004年1月11日
- 馬場基、「2004年全国出土の木簡」、木簡学会第26回研究集会、於奈良文化財研究所、2004年12月4日
- 山本崇、「2003年全国出土の木簡」、木簡学会第25回研究集会、於奈良文化財研究所、2003年12月7日