

平成23年 6月10日現在

機関番号：84604  
 研究種目：若手研究(B)  
 研究期間：2008～2010  
 課題番号：20720182  
 研究課題名(和文) 木簡の構文・文字表記パターンの解析・抽出研究  
 研究課題名(英文) Analysis and extraction study on syntax and character mark pattern of wood tablets  
 研究代表者  
 馬場 基 (BABA Hajime)  
 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 主任研究員  
 研究者番号：70332195

研究成果の概要(和文)：木簡積文へのタグ付きデータの作成を行った。タグ付きデータを『木簡画像データベース・木簡辞典』で公開しているデータと組み合わせることで、木簡画像データ上での記載内容の位置の把握を行うためのシステム開発を行った。この情報を利用して、記載内容木簡上位置情報による分析実験を行い、手法をほぼ確立した。

研究成果の概要(英文)：I put up tags to the text data of Wooden Tablets, and related that text data with tags to the image data of wooden tablets. By these studies, I developed the method for knowing the position of the content of the description on the wooden tablets, and succeeded in analyzing the wooden tablets.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008年度	2,000,000	600,000	2,600,000
2009年度	900,000	270,000	1,170,000
2010年度	500,000	150,000	650,000
年度			
年度			
総計	3,400,000	1,020,000	4,420,000

研究分野：日本史

科研費の分科・細目：史学・日本史

キーワード：タグ付け・出土文字資料・木簡

## 1. 研究開始当初の背景

出土文字資料を読む行為は、「経験と勘」に多く依拠しており、職人的な経験と技術の蓄積も欠かせない。だが同時に、経験を体系化して共有することで、より効率的で質の高い木簡積読が可能になると考えた。

一方、科学研究費補助金「推論機能を有する木簡等出土文字資料自動認識システムの開発」(基盤S：研究代表者 渡辺晃宏。

研究分担者として参加)で開発された「木簡画像データベース・木簡辞典」等の技術により、木簡情報のデジタル化が飛躍的に進みつつあり、また理科系技術者との共同研究・交流の基盤も整ってきていた。

## 2. 研究の目的

木簡のフルテキストデータ・画像データを活用し、記載内容を整理したタグ付きデ

ータを作成する。

また、これまで学会で蓄積されてきている木簡に関する経験知（木簡の構文パターン、構文の木簡上での配置、木簡の形状・大きさとの関連性などの諸点）を整理する。

タグ付きデータとして整理された木簡データに基づき、上記の経験知を数値的に捕らえて検討して、その蓋然性についての知見を得、客観性の高い知に展開すると共に、これまで気づかれていない構文等の様相の発見を目指す。

### 3. 研究の方法

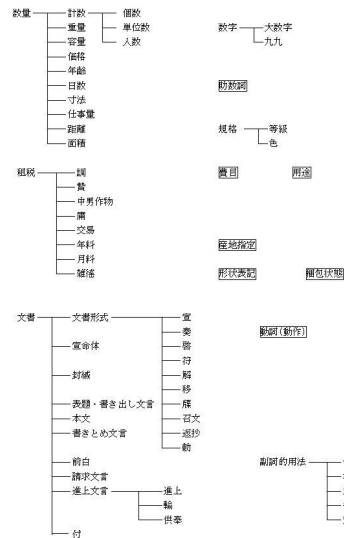
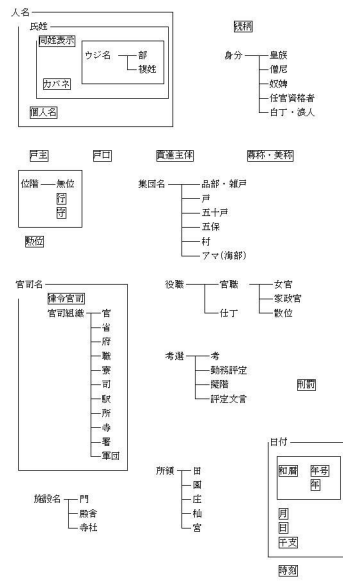
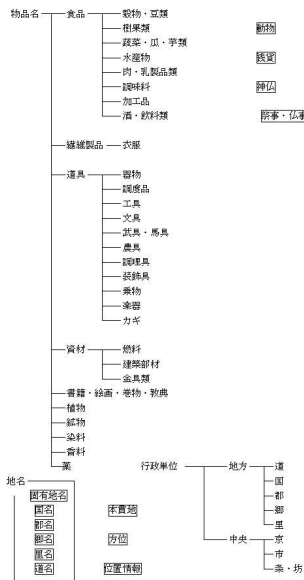
次の各作業をそれぞれ連続的に蓄積しながら研究を進めた。

- 1 既往の木簡構文研究の整理
- 2 既往の経験的知識の整理
- 3 木簡積文テキストデータへタグ付け
- 4 木簡簡画像上での文字の位置データ付与
- 5 データの対照・検証
- 6 実際の難読木簡での実験
- 7 新規の構文パターンの抽出と蓄積

特に、3の作業は本研究を推進する上で機軸となるデータの作成作業であり、また時間と手間が必要であることから、多くの時間を割いた。

### 4. 研究成果

- (1) 木簡積文へのタグ付きデータの作成  
木簡の積文へのタグの付与を行った。



この作業に際して、どのようなタグをどのような階層で付与するかを検討を事前に行い、また作業の進展に伴って発生した問題にも対応しながらタグの選定・構成を行った。こうした検討を踏まえ、タグは上記一覧表の内容で付与した

『平城宮木簡』1～6、『平城京木簡』1～3についてのタグ付けを完了した。この結果、現時点である程度評価が安定した奈良時代都城出土木簡についてのタグ付け作業がすべて出来たといえることができる。

このタグ付きデータは、「木簡画像データベース・木簡辞典」に活用し、タグ情報に基

づくリンクなどを可能にし、研究を加速するための研究資源として提供する予定である（タグ付きデータを利用した「木簡字典」は、サーバの調整のため、2011年6月現在奈良文化財研究所外未公表。所内で試験運用中。）

## (2) 木簡画像上での記載内容把握方法開発

タグ付きデータを、『木簡画像データベース・木簡字典』で公開しているデータと組み合わせることで、木簡画像データ上での記載内容の位置の把握を行うためのシステム開発を行った。

タグ付きデータを、『木簡字典』のもつテキストデータをもとに、画像データを参照する。そして、ソフトウェアによって、木簡全体のメタデータ（画像の大きさ含む）と共に、タグ付きのデータ、タグごとの画像上の位置情報などを書き込んだXMLファイルを作成して、蓄積するものである。

『木簡字典』で利用しているデータ・システムの特性から、ピクセル単位の位置情報ではなく、やや大きめの単位での位置情報であり、また画像という性格上、木簡写真の背景にある「余白」も含まれてしまっている。こうした点は、分析の精度を考える上で、問題になる可能性があった。ただし、分析作業における「遊び」の確保や、作業の効率などから、上述の問題点は軽微であろうと判断し、作業を進めた。なお、後述する実証実験ではその問題点についても検討し、多様な形態の木簡を俯瞰的に見渡し、誤差などを吸収することができるので、むしろ分析に優位であることが判明している。

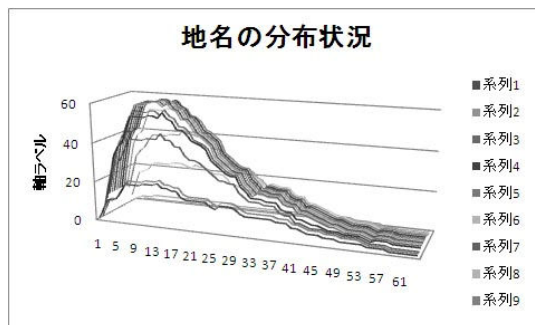
『木簡字典』でデータが公開されている範囲という限定が加わるものの、画像を媒介として木簡上のどの場所にどういった意味内容が書かれているかを、数値的に分析するための基礎的なデータの作成のための手法・手順を開発することに成功した。

これによって、今後データが増えても、基本的に同様の手法でデータの蓄積を進めていくことができる素地を確立することができ、今後木簡の文字情報・意味情報を画像情報や木簡そのものにおける位置と結びつけながら考えるための基盤が確立できたと考える。

## (3) 記載内容木簡上位置情報による分析実験

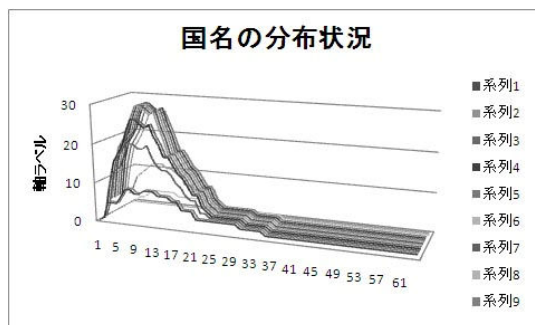
上記作業で得られた、記載内容の木簡上位置情報を用いた分析を行った。これによって、いくつかの方向性を得ることが出来、こうした一連の作業が木簡の研究・分析の上で有効であることが確認できた。以下、分析事例をいくつか紹介する。

地名のみでは、以下の分布になる。



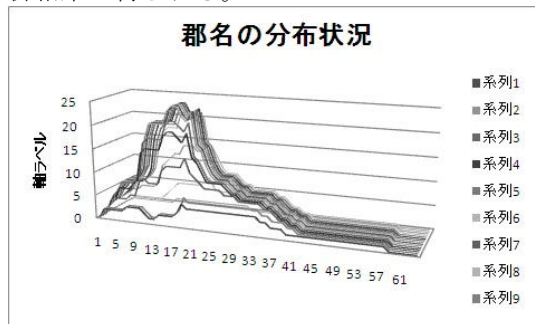
地名は全体としては木簡の上部半分に記載が集中する傾向にあるが、下部の記載が全く無いわけではない。

次に、地名でも、国名に絞って検討すると以下のような結果になる。



国名記載が、地名全体以上に極端に木簡の上部に集中することが知られる。逆に言えば、木簡の中央より下に国名がくることはまずないという様相が確認できる。

郡名について検討すると以下のような分析結果が得られる。

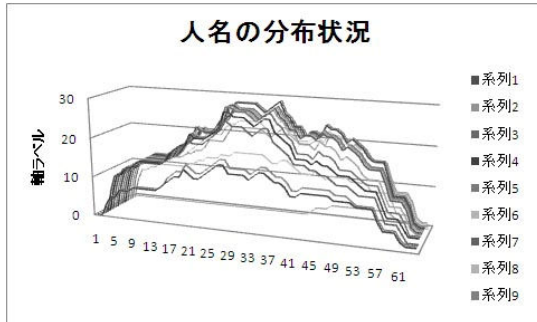


郡名は、国名とはやや異なった分布をみせる。若干、木簡の中央部付近にピークがずれ、木簡の下部に記載がされる場合もある。これは単に国名の後に郡名が記載される、というだけではなく、郡名の場合木簡の下部に記載されることもあるという事実も示している。

以上「地名」という大分類全体で、その記載場所に一定の傾向が見られることが分かると共に、その中に含まれる要素でも、それぞれ木簡上で記載される場所の傾向には差異が存在する傾向が明らかになった。

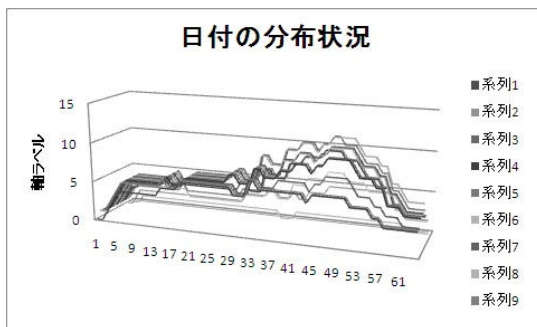
一方、「地名」以外の大分類の項目では当然地名と異なった分布が想定されるが、実際どうであろうか。人名について分析すると、

次の様な結果が得られた。



人名は、地名とは大分異なった分布を見せる。木簡の中央部付近にピークがくるものの、国名に見られた程の極端な偏りは確認できない。

さらに、年月日などの日付関連の記載を分析すると、次のようになる。



日付の記載については、木簡の後半部分への記載の偏りが確認できる。

なお、今回は分析していない7世紀代の木簡を分析すれば、また違った傾向が見られる可能性が考えられ、時期による変化も可視的・数値的に把握することができるであろう。

以上の実験から、木簡の記載内容の木簡上での記載位置を数値的に分析することは有意義であり、かつ新しい知見を生み出しえる可能性をもつことを確認できたといえる。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計13件)

- ①馬場基「荷札と荷物のかたるもの」『木簡研究』30、2008。233-264。査読有
- ②未代誠仁、齋藤恵、戸根康隆、石川正敏、中川正樹、馬場基、渡辺晃宏「古代木簡解読支援のための文字パターン検索」『情報処理学会論文誌』Vol.50(4)、2009。1444-1455。査読有
- ③未代誠仁、中川正樹、馬場基、渡辺晃宏「デジタルアーカイブ検索機能を有する木簡解読支援システム」『日本情報考古学会第26回大会講演論文集』Vol.6-2009、2009。p.p.66-77。査読有

- ④馬場基「古代下級官人出勤日数実態調査」『日本歴史』729、2009。88-89。査読有
- ⑤馬場基「平城京という「都市」の環境」『歴史評論』728、2010。46-61。査読有
- ⑥馬場基「木簡の世界」田辺征夫・佐藤信編『古代の都2 平城京の時代』吉川弘文館、2010。149-173。査読無
- ⑦馬場基「木簡の作法と1000年の理由」『日韓文化財論叢Ⅱ』独立行政法人国立文化財機構奈良文化財研究所・大韓民国国立文化財研究所、2010。pp365-379。査読無
- ⑧Sherini Somayah, 未代誠仁, 中川正樹, 馬場基, 渡辺晃宏「古代木簡解読支援システムにおける字体検索の高性能化」『人文科学とコンピュータシンポジウム論文集』2010-15、2010。pp.27-32。査読有
- ⑨馬場基「平城京」佐藤信編『史跡で読む日本の歴史4 奈良の都と地方社会』吉川弘文館2010。pp10-38。査読無
- ⑩馬場基「平城宮」佐藤信編『史跡で読む日本の歴史4 奈良の都と地方社会』吉川弘文館2010。pp39-60。査読無
- ⑪馬場基「奈良の寺々」佐藤信編『史跡で読む日本の歴史4 奈良の都と地方社会』吉川弘文館2010。pp61-94。査読無
- ⑫馬場基「寺院の建設」『季刊考古学』112、2010。pp34-37。査読無
- ⑬Akihito KITADAI, Sherini SOMAYEH, Masaki NAKAGAWA, Hajime BABA and Akihiro WATANABE「Non-linear Normalization of Damaged Characters for Search Refinement」Proc. 2nd China-Japan-Korea Joint Workshop on Pattern Recognition、1、2010。pp.185-188。査読有

[学会発表] (計2件)

- ①馬場基・井上聡「文字字形総合データベース作成の試み」、第2回人間文化研究情報資源共有化研究会、2009/7/16。国文学研究資料館。
- ②Akihito Kitadai, Masaki NAKAGAWA, Hajime Baba and Akihiro Watanabe「A Combining Method of Non-linear Normalization to Support Reading Damaged Character Pattern on Historical Documents」13th Conference of the International Graphonomics Society, 2009/9/14, Dijon, France

[図書] (計1件)

- ①馬場基『平城京に暮らす』吉川弘文館、2010。総ページ数237

[その他]

ホームページ等

「木簡ひろば」

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

6. 研究組織

(1) 研究代表者

馬場 基 (BABA Hajime)

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・主任研究員

研究者番号：70332195

(2) 研究分担者

( )

研究者番号：

(3) 連携研究者

( )

研究者番号：