

令和 6 年 6 月 12 日現在

機関番号：34416

研究種目：基盤研究(C)（一般）

研究期間：2019～2023

課題番号：19K01044

研究課題名（和文）中国出土木簡の書体分析の基礎的研究

研究課題名（英文）A Basic Study on the Analysis of Writing Style of Excavated Wooden slips from China

研究代表者

藤田 高夫（FUJITA, Takao）

関西大学・文学部・教授

研究者番号：90298836

交付決定額（研究期間全体）：（直接経費） 3,400,000円

研究成果の概要（和文）：本研究では、木簡に記された中国古代の行政文書上の書体分析の手法開発を目指した。木簡の書体を分析するためには、一文字ごとの画像データの集積がまず必要となるが、本研究では中国木簡の赤外線写真から、AIを活用して文字を自動的に切り出すシステムの開発を試みた。具体的には、文字が独立して書写される秦漢木簡の特性に着目し、文字間の空白部分を鍵として文字境界領域を認識するアルゴリズムを実装化した。これと併せて、書体のくずれ度合いを定量的に判定する手法の開発に着手した。ここでは、くずれた文字と謹直な文字の2つの文字群を準備し、深層生成モデルを学習させて差分を計算することでくずれ度合いの算出が可能となった。

研究成果の学術的意義や社会的意義

本研究の学術的意義は、従来ほとんど着目されなかった木簡の書体（書きぶり）を対象として研究するための基盤を提供したことにある。これまでの木簡研究は歴史史料としての価値のみが偏重され、出土物としての側面が等閑視されてきた。そのため、テキスト化された記述内容のみの分析で事足りるとされ、木簡の形状や文字の位置など、木簡が持つ豊富な情報が捨象されてきた。本研究は、文字切り出しシステムの開発によって大量の文字画像の集積を極めて容易なものとし、これまで着目されなかった木簡の書体分析という未開拓の研究分野にアクセスする上での最大の障壁を取り除き、定量化されたデータに基づく書体分析の道を開いた。

研究成果の概要（英文）：This study aimed to develop a method for analyzing writing style on ancient Chinese documents written on wooden slips. In order to analyze the writing style, it is first necessary to accumulate image data for each character. In this study, we attempted to develop a system to automatically cut out characters from infra-red photographs of Chinese wooden slips using AI. Specifically, we focused on the characteristics of the Qin-Han wooded slips, in which the characters are written independently, and implemented an algorithm that recognizes character boundary areas using the blank spaces between characters as a key. In addition to this, we also started to develop a method to quantitatively determine the degree of writing distortion. The degree of distortion can be calculated by training a deep generative model and calculating the difference between two groups of characters, one with distortions and the other with respectability.

研究分野：中国史

キーワード：木簡 居延漢簡 書体分析 深層学習 文字切り出し 文字のくずれ度合い 同筆と異筆 AI

1. 研究開始当初の背景

- (1) 中国秦漢時代の一次史料である木簡が最初に発見されてからすでに 1 世紀以上が経過した。中国古代木簡の大多数は、日々の行政処理において生み出される文書や記録類であり、その記述内容は文献史料ではうかがえない文書行政の実態を詳細に伝えるものであった。しかしながらこれまでの木簡研究は、歴史史料としての活用に重点が置かれ、テキスト化され活字として出版された記述内容を理解すれば事たれりという研究が一般的で、木簡の形状や材質、文字の書かれる位置など、出土史料である木簡が本来含有している様々な歴史情報が捨象されてきた。その中でも、書体すなわち文字の書きぶりの違いが、文書行政においてどのような意味を持つのかは、研究者の関心外であった。
- (2) 中国木簡の書体の分析がこれまで等閑視されてきた背景として、まずデータの膨大さがある。現在 10 万件を超えた中国出土木簡に含まれる文字数は、おそらく 100 万のオーダーであり、分析のための画像データ蓄積に多大の労力と時間がかかったためである。また素材となる画像(木簡の写真)も発掘報告書の年代が古いものであると極めて不鮮明で、画像解析の対象となり得なかったことも、書体分析が進まなかった一因である。しかし近年の ICT の発達は、手作業による文字データ蓄積を自動化する可能性を拓いた。また 21 世紀になると古い報告書の再撮影も含めて大量の赤外線写真の公開が進み、これによって簡の汚れが除去され墨跡のみの精細な画像の利用が可能となった。これらの状況の変化によって、木簡の書体分析を阻んできた大きな阻害要因が除かれ、書体分析の手法確立に向けての条件が整った。

2. 研究の目的

- (1) 本研究の最終的目的地は、字体を中心に木簡文書の書きぶりに焦点を当てて、漢代の文書行政における個々の文書の位置づけを再吟味する新しい視角を提示することである。例えば右に示すように、「書」という 1 文字をとりあげても大きな幅があるが、謹直な隷書を用いた文書と崩れた草書を用いた文書では、記述内容が同じであっても、文書としての性格が同一とは考えがたい。本研究は、このような字体の差異と文書の性格との相関関係に光を当てるものである。
- (2) 本研究では書体分析のための基礎資料となる 1 文字ずつの木簡画像の集積を進めることになる。そのためには、上述の阻害要因の一つである文字切り出しの自動化は、研究遂行上の必須のプロセスである。そうして集積された画像データは、汎用的な「木簡字典」としての機能を有するものとなり、木簡研究者のみならず、手書き文書を扱う歴史研究者や文字学者に多大な裨益をもたらすことが期待される。こうした切り出しシステムが実用化されれば、文字データの集積が極めて容易になり、広く古文書学にも書体分析という研究領域を開放することにつながるであろう。



3. 研究の方法

本研究は、以下のような手順で遂行した。

- (1) 木簡のデジタル画像の集積
漢代木簡の図版をスキャンし、木簡一件ごとに JPEG 画像として切り出す。対象は鮮明な赤外線写真の存在する「肩水金關漢簡」「地湾漢簡」「敦煌漢簡」「台湾中央研究院所蔵居延漢簡」など約 2 万件を超えるデータとなる。
- (2) 文字画像の切り出し
木簡から文字を一つずつ切り出す。この作業は極めて手間のかかるものであるが、AI を活用した文字切りだしツールを実装し、作業の効率化を図る。さらに文書に頻見する文字(「年」「府」「官」「尉」「書」「言」など)を中心にデータを集積する。
- (3) 文字の謹直度レベルの設定
切り出した文字全体を俯瞰して、謹直度レベルを 5 段階で設定する。謹直度判定に客観性を持たせるため、漢碑の隷書を最謹直とし、縦画の垂直度、横画の水平度、筆画の等間隔性、波磔の独立性などを指標とする。これらのデータをもとに、AI による文字の謹直度認識のアルゴリズムを創出する。
- (4) 簡牘全体の謹直度レベルの算定：上記(2)の文字を複数種ふくむ簡牘を対象に、(3)で設定した各文字のレベル値を積算・平均し、1 件の簡牘の謹直度レベルとして付与する。

4. 研究成果

- (1) 報告書等から木簡 1 件ずつの画像を切り出す作業は、当初予定していた「肩水金關漢簡」「地湾漢簡」「敦煌漢簡」「台湾中央研究院所蔵居延漢簡」に、研究期間中に刊行が開始された「懸泉漢簡」を加え、総計で 4 万ほどのデータを集積した。JPEG 画像を選択したのは、

研究者の日常的な作業効率上 TIFF ではスペック的に苦しいと判断したからである。

- (2) 木簡画像から一文字ずつを切り出す作業の自動化は最も困難なステップであった。秦漢木簡は、一文字ずつが独立して書写され、複数文字が連続するいわゆる「つづけ字」は基本的に存在しない。この特徴に着目して本研究では、木簡文字の書写領域に加え、文字と文字間の空白部分を認識することによって、文字領域の認識精度を格段に向上させることに成功した。右の図1に示すように、(a)の木簡写真から(b)のように一文字ずつを切り出すことが高い精度で可能となった。その際に(c)のように墨跡のある領域に加え、(d)のように文字間の空白領域を認識することによって、文字切り出しの精度を上げることができた。その結果、(a)のような、ほぼ同じ大きさの文字が比較的等間隔に配置された木簡については、100%に近い精度で、文字を認識することができている。ただし、木簡の中には、右に示した木簡のように一件の簡に異なった大きさの文字が混在する

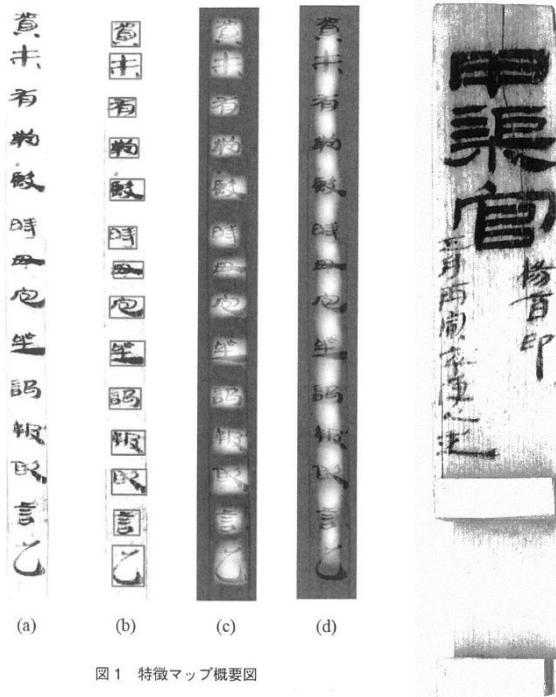


図1 特徴マップ概要図

例も多数あり、この場合には文字切り出しの精度がそのままでは大幅に低下することも判明した。そのような木簡に対しては、文字間の空白の識値を調整することで都度対応し、正しい文字切り出しデータを抽出した。このようなタイプの木簡をAIが正しく認識するように訓練するには、また別の工夫が必要であり、これは今後の課題である。現時点で、文字切り出しの精度は、全体ではほぼ90%程度であるが、書体分析のためのデータ集積という観点からは、十分に実用に耐えうるレベルに達している。

- (3) 文字の謹直度レベルの設定、すなわち文字のくずれ度合いの定量的把握については、帰順となる謹直な隷書体からの乖離度を測ることを試みた。具体的には、隷書から草書にいたるくずれを0(謹直)から1(くずれ)の間の連続値で定義される謹直度を推定する方法を構築した。その基礎となるのが、深層画像生成法の一つである敵対的生成ネットワーク(Generative Adversarial Network: GAN)である。GANは、訓練データに存在しないパターンを再現して画像生成することができない。この性質を利用して、正常画像を用いて学習されたGANを用いて異常画像を再構成(入力画像と似た画像を生成)しようとした際の誤差から正常・異常を判定することができる。本研究では、隷書と草書の文字画像をGANに入力した際の生成結果の比較を行った。図2は「包」の例を表しており、上段が入力文字、左端が隷書、それ以外は

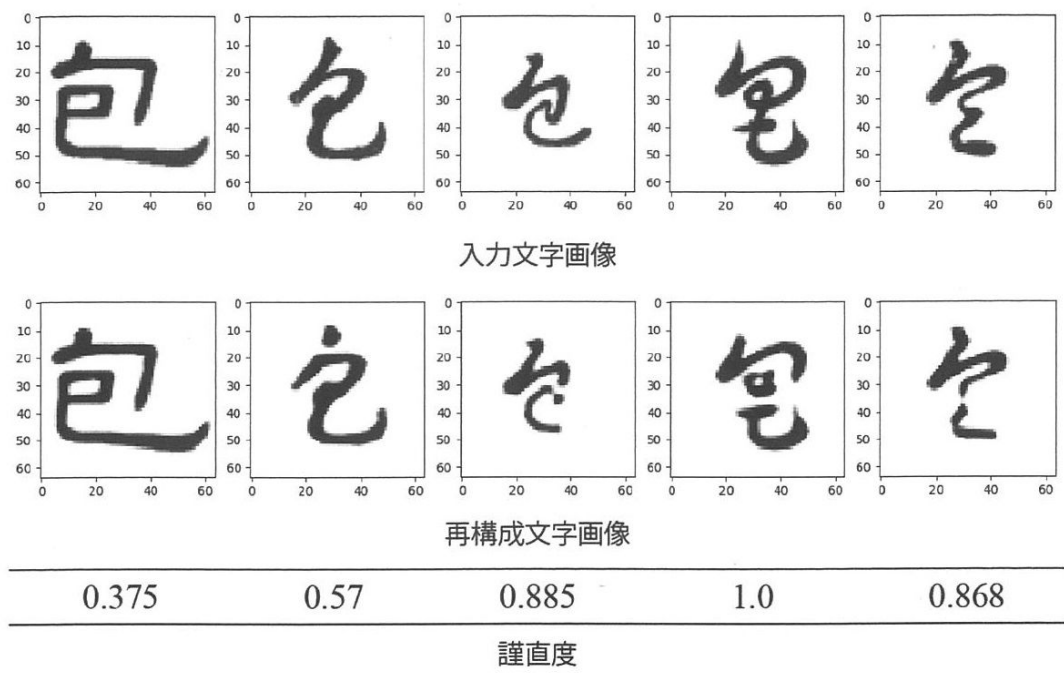


図2 謹直度判定例

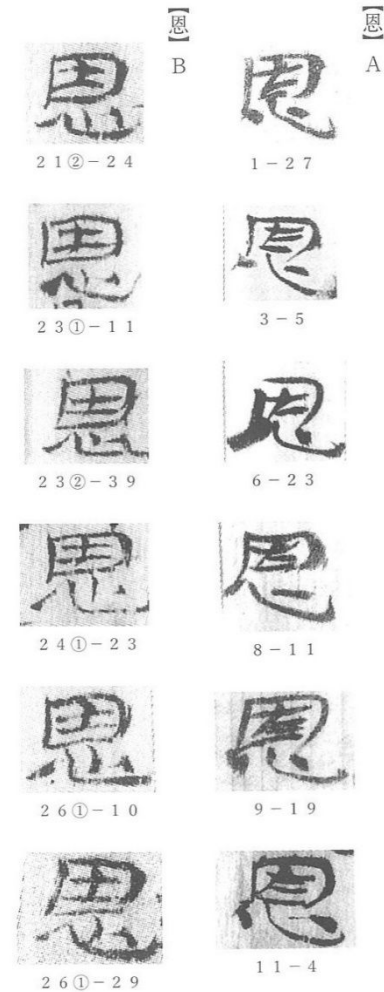
草書を表す。中段は、上段の文字からそれらと最もよく似た文字を GAN が再構成した結果である。つまり、草書のくずれに応じて、GAN の入力文字と再構成文字との間の誤差が大きくなることが確認される。このように謹直度の大小が入力文字のくずれと相関があることが判明し、ここから文字の謹直度を客観的に提示することが可能となった。

- (3) 本研究が木簡の書体分析手法の確立を目指す過程で浮上してきたのが、同筆と異筆の問題である。多くの木簡は単独簡として使用されるのではなく、元来は紐で編綴された「冊書」として存在していた。それが発掘された時点では紐は失われバラバラになって出土する。したがって、冊書の復原は木簡研究の終着点の一つでもあるが、その復原の際に有力な基準となるのが筆跡の同一性である。本研究が蓄積する文字画像データは、よく似た書きぶりの文字の検出も可能とするものであり、それによって新たな冊書復原の可能性を拓くものとなる。今回の研究課題そのものではないが、同筆と異筆についても書き手の違いによって文字がどれくらい変化するか、さらに同一の人物が同じ文字を書いた際に書きぶりにどれくらいの幅があるかについても、文字画像データを利用して基礎的な考察を行った。右に示したのは、ある裁判案件についての 2 件の報告書から抽出した「恩」字である。文書 A と文書 B では、明らかに書きぶりが異なり、両者の書き手が別であることは明白である。同時に、それぞれの文書で複数回出現する「恩」字の書きぶりの特徴は、文書 A、B それぞれで一貫していることも見て取れる。このような事例を集積していくことで、同筆簡を集成し冊書復原につなげていくことが可能となるのも、本研究の付随的成果の 1 つである。

引用文献

藤田高夫、吉田壮、江孟原、「中国西北辺境出土木簡の書体分析の基礎的研究」、『考古学ジャーナル』2021 年 10 月号、p.33 ~ p.36

藤田高夫、「漢代木簡の同筆判定のための基礎的考察」、『関西大学東西学術研究所創立 70 周年記念論文集』、関西大学東西学術研究所、2022 年、p.357 ~ p.376



5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 2件/うち国際共著 0件/うちオープンアクセス 2件）

1. 著者名 H. Shimoyama, S. Yoshida, T. Fujita, and M. Muneyasu	4. 巻 vol. E106-A, no. 11
2. 論文標題 U-Net Architecture for Ancient Handwritten Chinese Character Detection in Han Dynasty Wooden Slips	5. 発行年 2023年
3. 雑誌名 IEICE Trans. Fundamentals	6. 最初と最後の頁 1406-1415
掲載論文のDOI（デジタルオブジェクト識別子） 10.1587/transfun.2023SMP0007	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 藤田高夫	4. 巻 14
2. 論文標題 漢代の部都尉をめぐる覚書 中国王朝の辺境政策の一側面として	5. 発行年 2023年
3. 雑誌名 東西学術研究所研究叢刊	6. 最初と最後の頁 1-24
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 藤田高夫	4. 巻 -
2. 論文標題 書体分析による中国木簡学の可能性	5. 発行年 2022年
3. 雑誌名 KU-ORCASが開くデジタル化時代の東アジア文化研究	6. 最初と最後の頁 11-21
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 藤田高夫	4. 巻 -
2. 論文標題 漢代木簡の同筆判定のための基礎的考察	5. 発行年 2022年
3. 雑誌名 関西大学東西学術研究所創立70周年記念論文集	6. 最初と最後の頁 357-376
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 藤田高夫、吉田壮、江孟原	4. 巻 2021-10
2. 論文標題 中国西北出土木簡の書体分析の基礎的研究	5. 発行年 2021年
3. 雑誌名 考古学ジャーナル	6. 最初と最後の頁 33-36
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計9件 (うち招待講演 1件 / うち国際学会 2件)

1. 発表者名 Takao FUJITA
2. 発表標題 New Developments in the Study of Ancient Chinese History by Excavated Materials
3. 学会等名 Kansai University & KU Leuven Joint Workshop on Historiography (招待講演) (国際学会)
4. 発表年 2024年

1. 発表者名 藤田高夫
2. 発表標題 中国木簡筆跡研究の課題と展望
3. 学会等名 関西大学アジア・オープン・リサーチセンター第2回研究例会
4. 発表年 2022年

1. 発表者名 下山 萌純, 吉田 壮, 藤田 高夫, 棟安 実治
2. 発表標題 U-Netに基づく文字の境界に注目した中国酒度木簡からの文字領域検出手法
3. 学会等名 電子情報通信学会技術報告, スマートインフォメディアシステム研究会
4. 発表年 2022年

1. 発表者名 江 孟原, 吉田 壮, 藤田 高夫, 棟安 実治
2. 発表標題 深層生成モデルを用いた中国出土木簡の書体分析に関する一検討
3. 学会等名 電子情報通信学会技術報告, スマートインフォメディアシステム研究会
4. 発表年 2021年

1. 発表者名 藤田高夫
2. 発表標題 漢代の部都尉をめぐる問題 中国王朝の辺境政策の一側面
3. 学会等名 関西大学東西学術研究所創立七〇周年記念シンポジウム
4. 発表年 2021年

1. 発表者名 Takao FUJITA
2. 発表標題 Analysis of Writing Styles on wood slips of the Han period
3. 学会等名 関西大学オープン・リサーチセンター国際シンポジウム
4. 発表年 2019年

1. 発表者名 藤田高夫
2. 発表標題 An Attempt to Analyze Writing Forms in Administrative Documents and Records in the Han Period
3. 学会等名 東アジア文化交渉学会第11回年次大会(国際学会)
4. 発表年 2019年

1. 発表者名 吉田 壮, 谷知 真行, 藤田 高夫, 棟安 実治
2. 発表標題 深層学習を用いた中国漢代木簡の文字領域検出
3. 学会等名 電子情報通信学会技術報告, スマートインフォメディアシステム研究会
4. 発表年 2020年

1. 発表者名 Soh Yoshida
2. 発表標題 Image analysis for character region extraction from wood slips
3. 学会等名 The 3rd International Symposium at KU-ORCAS
4. 発表年 2019年

〔図書〕 計1件

1. 著者名 TAO Demin, FUJITA Takao	4. 発行年 2021年
2. 出版社 V&R unipress	5. 総ページ数 323
3. 書名 Cultural Interaction Studies in East Asia; New Methods and Perspectives	

〔産業財産権〕

〔その他〕

KU-ORCAS 木簡学の魅力(筆文字をデジタルの目で) https://www.ku-orcas.kansai-u.ac.jp/digest/index.html

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分 担 者	吉田 壮 (YOSHIDA Soh) (70780584)	関西大学・システム理工学部・准教授 (34416)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関