

# 歴史的文字に関する既存知の集積と分析

高田祐一（奈良文化財研究所）

**概要：**歴史的な文字に関する既存知は、分散的・個別的に存在している。言語化されている場合でも、用語は各研究者によって異なり、統一がとれているとはいえない。一方で、各研究者が蓄積した知識には、文字に潜む歴史性や、文字から歴史を構築する際に貴重な情報となる内容も多く含まれている。これらを、解析・分析することを試み、その成果の一端と今後の展望について述べる。

**キーワード：**歴史的な文字、既存知、木簡、自然言語処理

## A Study of Accumulating and Analyzing Information related to Previous Studies of Characters Inscribed on Wooden Tablets

Yuichi TAKATA (Nara National Research Institute for Cultural Properties)

**Abstract:** Information of previous studies about characters inscribed on wooden tablets exists individually. Although these data are published, the way of using technical terms differs depending on each researcher, thus the format of written information is not taking unified. However, this existing information includes important knowledge to understand Japanese history that derives from characters written in historical records. Therefore, this study tries to analyze the distributed information by the natural language processing system, and consider its prospect for the future.

**Keywords:** character, existing information, wooden tablets, natural language processing system

### 1. 木簡の調査研究の概観

文献史学・考古学は、蓄積型の学問であり、研究史が重要である。調査事例や研究成果を蓄積していくことで、研究が広がり深化していく。蓄積こそが重要であると同時に情報量が膨大になっている現代において、どう蓄積しどう有効に活用していくかが問われている。

調査研究に携わる者が歴史的な文字を読み、蓄積した知識には、文字に潜む歴史性や、文字から歴史を構築する際に貴重な情報となる内容も多く含まれている。しかし、歴史的な文字に関する既存知は、分散的・個別的に存在している。言語化されている場合でも、用語は各研究者によって異なり、統一がとれているとはいえない。本稿では歴史的な文字に関する知識を集積し、それらを、解析・分析することを試み、その成果の一端と今後の展望について述べる。

歴史的な文字に関する知識を研究資源化し、既存知を集積し活用できれば、奈良文化財研究所（以下、奈文研）の木簡整理調査の研究業務の水準の維持、高次化に貢献できる。

本稿で扱う対象は、木簡の歴史的な文字とする。分析方法は、木簡に関する報告書・論文から歴史的な文字に関する記述を抜き出し、データ化する。データクレンジングのうえ、自然言語統計処理の観点から分析・解釈を行うものである。

### 2. 木簡の調査研究の概観

奈文研では、発掘調査で出土した木簡を「洗浄する→写真撮る→文字を読み取る→報告書を作る」といった業務



図1 平城宮跡 SK820 発掘調査風景（1963年8月）



図2 木簡洗い風景

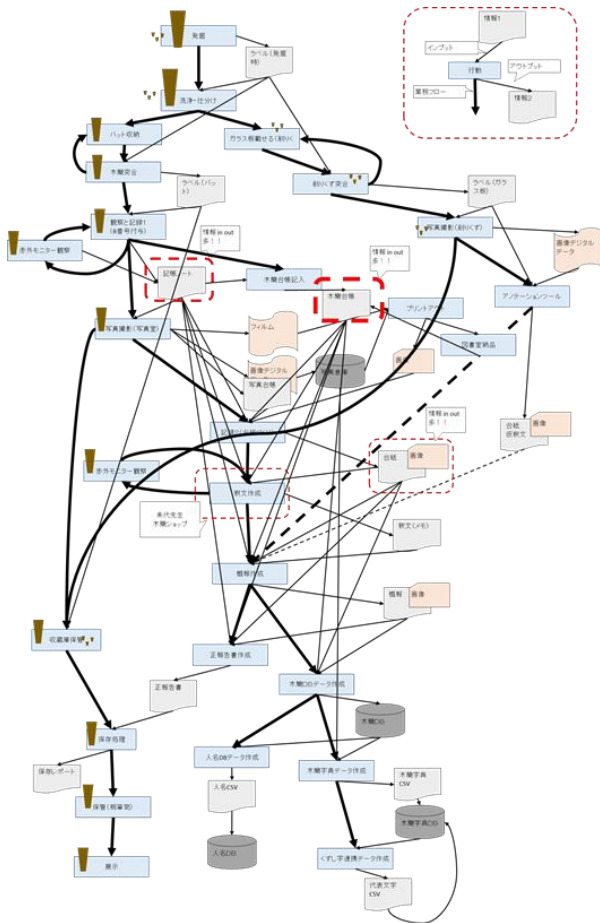


図3 木簡関係業務全体プロセス図

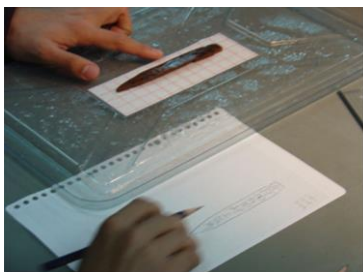


図4 木簡の観察結果を記載する作業風景

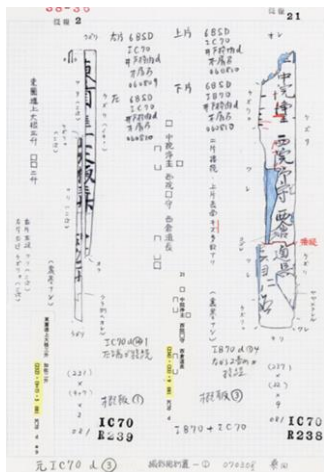


図5 木簡の観察結果を記載する記帳ノート

がある。具体的には、発掘調査で出土した木簡を土と木簡に分離し洗浄する工程(図1・2), 出土木簡を個別にパッケージしていく整理工程, 写真撮影の工程, 木簡の文字・内容・形状などを観察する記帳・釈読工程(図4・5), 過去の研究史などを整理し具体的成果をまとめる報告書作成工程, 公開データベース用のデータ作成工程などである。出土からデータベース作成まで膨大な手続きを経て情報を整理し, 各工程では情報のインプット・アウトプットを繰り返す(図3)。これらの工程のうち, 報告書作成やデータベース作成は, 学術成果として報告書やデータベースとして見える形の具体的なモノという形式知として扱え, 社会に還元される。しかし, 洗浄工程や釈読工程では, 作業をしながらノウハウや気づき, 仮説が発生しているが, 特にデータ化しないため, 霧散している状態である。

### 3. 木簡において何を認知しているか

#### 3.1 木簡調査の工程において観察ポイントは異なるか

奈文研では「気づきメモ」という文字化を行うことによって, 霧散していた情報の可視化を推進している。

「気づきメモ」とは, 洗い～釈読～報告書作成等の各工程に従事する職員が自由記述で, 気づいたことを書いて貯めたメモである。原則, 1項目につき1枚のメモに記載し, 長文によって複数要素を盛り込むことは禁止である。内容に制限はなく, 「AとBは似ている(かもしれない)」といった推測や主観など何でもよい。また「木簡の切り込みがこういう形の場合は, ここを持てば扱いやすい」といった経験による作業ノウハウ等も対象とする。まずは気づきを多数蓄積するのが主眼である。

試行的に取り組んでいる「気づきメモ」300個(文字数12,853字)と研究史メモ(後述)を対象に行った, KHcoderによる計量テキスト分析の例を紹介する。

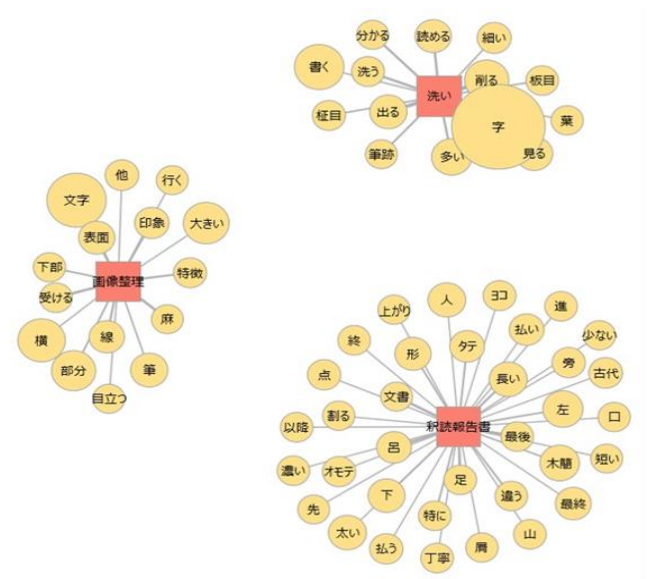


図6 工程別気づきメモの語句の抽出語

まず、作業工程を①木簡の洗い工程②画像整理工程③木簡の積読や報告書作成工程の3つに分け、多く出現した語を確認し、語の出現順位を整理した。

①の洗い工程では、「削る」「柁目」「板目」「木片」などモノそのものに関する気づきが多い。また「出る」「読める」「分かる」など、木簡への作業者の認識や身体動作に関する語も多い。この工程は泥がついている木片から文字を探す作業となることから、文字があるかないか、読めるかということに意識が集中するため「出る」「読める」「分かる」といった語になったと推測される。

②の画像整理工程では、「印象」「表面」「裏面」「大きい」「小さい」など面的な認識に関する気づきが多い。この工程は画像を整理するために木簡の全体を把握する必要があることから、面的な語になったと推測される。

③の積読や報告書作成工程では、「横」「縦」「部分」「右」「左」「長い」「太い」など、個別の部分に関する気づきが多い。この工程は積読するにあたり、丁寧に1字ずつ観察するため、細かい部分に着目する語が多くなったと推測される。

以上のように、工程によって、語の出現傾向に違いが見られた。これは、作業者の意識の違いに起因すると考えられる。さらに工程別に抽出語のバリエーションを図化した(図6)。

①の洗い工程に比べ、③の積読や報告書作成工程では語が分散する傾向にある。洗い工程では、字があるかないかという点に作業者の意識が集中するが、積読や報告書作成工程では多様な文字のクセなどを観察するため、様々な部分について考慮する必要があり、その結果が「気づきメモ」にも表れていると考えられる。積読では、様々なことを観察し、考えながら従事しており、発生している情報量が多い工程であるとわかる。

### 3.2 研究者は木簡の何を観察しているか

既に発行されている論文等にて、木簡の文字に関する言及を抜き出し、1要素1行にして1758個リスト化した(言及の説明文字数 66920。以下、研究史データ。)。データクレンジングとして、参考文献などは除外し文字に関する言及のみとした。このリストに対し、KH Coder(Version:2.00f[Perl 5.14.2,Perl/Tk804.029])にて、品詞別

項番	木簡ID	記載日	記載者	観点分類	内容
1	1	2015/1/26	方	木簡の形	【市大樹「飛鳥藤原木簡の研究」P363】個別に完形木簡をみると、尾張国では平均長さ126mm幅27mmで太く強いものが多い。
2	1	2015/1/26	方	木簡の形	【市大樹「飛鳥藤原木簡の研究」P363】尾張国では平均長さ180mm幅27mmと全体的に大型であるが、とりわけ長大なものが多い。
3	1	2015/1/26	方	木簡の形	【市大樹「飛鳥藤原木簡の研究」P363】美濃国では平均長さ160mm幅26mmであり、尾張国ほどではないがやや大型のものが目立つ。
4	1	2015/1/26	方	木簡の形	【市大樹「飛鳥藤原木簡の研究」P363】夢野国では平均長さ159mm幅23mmであり、長さの割にはやや幅が狭い。
5	1	2015/1/26	方	木簡の形	【市大樹「飛鳥藤原木簡の研究」P363】若狭国では平均長さ154mm幅26mmで平均値とほぼ相似形をなすが幾分か大きいといえる。
6	2	2015/1/26	方	形式(書式)	【市大樹「飛鳥藤原木簡の研究」P367】二行書きの大半は片面使用であり、藤原・若狭国に大きく偏っている。

図7 研究史のテキスト分析

の抽出語を出力した。研究者が文字についてどう表現したかを分析するため、形容詞について出現回数順に出力した(表1)。

表現として、大きい-小さい、濃い-薄い、太い-細いであった。実際の言及を確認すると、文字が大きい/小さい、墨が濃い/淡い、字が太い/細いという用法であった。観察のポイントはある程度似ており、類型化できよう。しかし、多い/珍しい/少ないなどは、深いドメイン知識が必要となる。多いというのは、何を基準にしてどこからが多いという感覚か、経験知を明らかにする必要がある。この研究者個々が持っている何らかの基準を形式知化することで、既存知にアクセスができるようになる。また整理の過程で、新たな研究の気づきも発見できるだろう。

### 3.3 文字「国」の観察例

研究史データにおいて、国の文字について詳細確認した。研究史データの「国」の実際の使われ方を確認すると2パターンあった。1つ目は、国名に意味を持つパターンである。例えば「美濃国では平均長さ160mm幅25mmであり、尾張国ほどではないが、やや大型のものが目立つ」というように、木簡の特徴が国毎に異なるという内容である。収集した研究データという制限があるものの、過去の研究史においては、木簡の比較検討の際、国毎に差異を見出すことが大きな研究視点のひとつであったということであろう。2つ目は、「国」自体の形に注目した研究である。例えば「木簡

表1 研究史データの形容詞抽出語

形容詞	出現回数
多い	44
大きい	41
濃い	36
淡い	35
薄い	35
小さい	25
長い	24
高い	21
近い	19
太い	15
珍しい	15
重い	11
難しい	11
細い	10
悪い	9
強い	9

にはあまり見られない国構えの中が「玉」となっている「国」字形などである。このように、研究史を整理するうえで、定量的にテキストを把握できることは、研究者がよりよくデータを理解することの助けとなるだろう[1]。

## 4. 今後の展望

### 4.1 既存データベースへの研究史データの連携

奈文研が運用している木簡画像データベース「木簡字典」

(<http://jiten.nabunken.go.jp/index.html>)に、先述の文字に対する研究史の言及を紐づけすることで、歴史的文字に関する情報基盤とすることができる。情報を統合することで、研究史を踏まえた漏れのない研究が可能となる。

## 4.2 シソーラスの必要性

古代の用語は、表記ゆれが存在する。表記ゆれはデータベースでの文字検索の際に網羅的な検索ができないという課題となる。例えば魚のイワシであれば、鰯・伊和志・伊和之・伊委之のように木簡で記述される場合がある[2]。異体字も存在する。このような表記ゆれは、すべてのパターンを把握している研究者であれば、個別に対応できるが、初学者では対応が困難である。木簡字典では、意味検索機能として一部の語に意味タグを付与し、データベースに実装済みである。今後は、言葉の階層性を加味したうえで、網羅的に整理し、シソーラスとして体系的に構築する必要がある。シソーラスを実装したデータベースにできれば、文字の払い・はらいなどの表記ゆれもまとめて検索できるようになる。

## 5. おわりに

文字に潜む歴史性や、文字から歴史を構築するために歴史的な文字に関する既存知を集積し分析する取り組みの一端を述べた。既存知を把握するために、研究者がどういった観点で木簡を観察し、観察結果を言語化しているか分析した。従来、研究成果は紙印刷物の報告書・論文であるが、本稿では展望として既存のデータベースに紐づけシソーラス構築の必要性を示した。

**謝辞** 本研究の一部は、JSPS 科研費 26244041・16H05881 の助成によるものである。

## 参考文献

- [1] 樋口耕一:社会調査のための計量テキスト分析 内容分析の継承と発展を目指して,株式会社ナカニシヤ出版,(2014).
- [2] 方国花:いわゆる国字について—木簡の用例を中心に—,古代の文字文化<古代文学と隣接諸学4>, p. 99-126,(2017).