

敦煌漢簡画像データベース

6 R-2

*上島 紳一 **園田 浩一 *大庭 脩

*関西大学文学部

**姫路獨協大学一般教育部

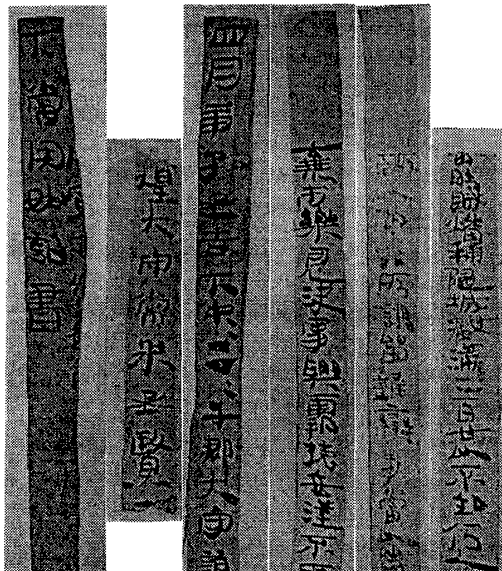
1. はじめに

中国漢代の一等資料である木簡は漢簡と呼ばれる。漢簡は実物が研究者の手元にないことが多く、研究著録は中国や台湾をはじめとする諸外国で出版された古い写真本に依存している。加えて、漢簡は地中での保存状態が悪い上、隸書体の文字は多様性に富み、文字や内容の釈読が研究者間で統一されていない簡が多い。

また、写真本に掲載されている漢簡写真の配列は研究者にとって必ずしも扱い易いものではない。漢簡研究の効率をあげる為には写真本から画像をデータ入力し、二次資料である釈読と同時に検索できるようにしておくことが望ましい。ここでは実験的な試みとして木簡の中から敦煌漢簡約千本を対象としてパソコンベースで研究者支援用の画像データベースを構築したので、その構成について述べ、簡牘研究への可能性について触れる。

2. 敦煌漢簡

木簡は紙が使えなかった頃、ヤナギ類の木に墨字で文字を書き、手紙や書類の代わりとして用いたものである(図1)。一本の木簡に書けないような内容の場合は数本の木簡の両端を紐でつなぎ、冊書にしてそこに文字を書いて用いていた。統一的な釈読の統一が取られていない敦煌漢簡が写真本の中の951本のうち、366本存在する。

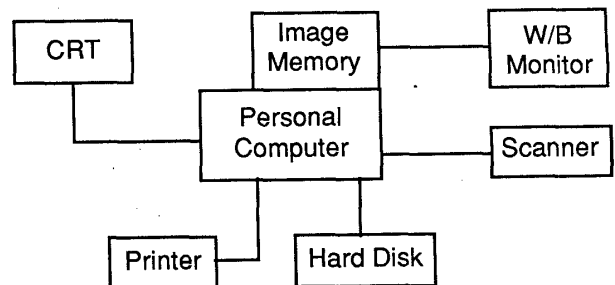


(図1) 敦煌漢簡冊書の一部

3. システム構成

本システムでは、利用者がほぼ文化系の研究者であるため、計算機に対して不慣れであっても容易に扱える必要があった。また、漢簡データは歴史的資料である為にデータの更新は頻繁でなく、保守作業の必要がほとんどないと考えられるので、研究者自身がシステムを個人的に保有して利用することが可能である。このような背景を考慮して本システムではパーソナルコンピュータを用いる。画像の利用できるデータベースソフトウェアもあるが、漢簡画像には判読が容易でないものもありパソコンの画像表示能力だけでは不十分である。そのため、データベースには漢簡の釈読に関するデータのみを登録し、漢簡画像は、本体に接続された 768*512 画素 8 ビットのイメージメモリを使用し、文字によるデータベースの検索に基づいて画像表示用モニタに高解像度で表示する。

画像データの入力には CCD カメラ とイメージスキャナにより行っている。撮影画像は、敦煌漢簡研究の基礎となっている 2 冊の写真本(1, 2)をページ単位に入力した上、一葉毎に切り出してファイル化している(図2)。



(図2) システム構成

ファイル構成は通常の文字コードを入力する為のデータベースと漢簡一葉毎の画像ファイルから構成されており、漢簡の二次資料はデータベースに漢簡一葉毎、共通理解の得られた釈文、官名、人名、地名、出土地、成語、単語、長さ、幅、行数などを入力している。二次資料データベースとしては文字のよる基本検索やそれぞれにテキスト研究においてよく知られているキーワードによる KWIC 整列化(Key Word In Context)が可能となっている(図3)。文字データベースでの検索や KWIC 整列化は、検

Image Database of Touen-Houang HAN Wooden Strips

*Shinichi UESHIMA, **Koichi SONODA, *Osamu OBA

*KANSAI University **HIMEJI DOKKO University

索結果を一旦ファイルに落とし、改めてエディタに取り込んで表示し、原簡番号により漢簡が表示される。画像ファイルとデータベースファイルを結合する為のインターフェースは C 言語を用いた外部プログラムで実現されている。

西大煎候嬰國下厥胡守士吏方承	書 從事下當用／者如律令	／令史	19
幼君少平足下吾母恙甚苦事尋幸賜	書 前請仲侯能軍叩頭辛甚甲午日朔		42
□□□□長寫移	書 到趣實籍部中移口		121
請輸札兩行隆五十繩廿支須寫下詔	書 到所移神爵四年十月盡五鳳元年五		137
門都尉子光丞萬年謂大煎都候寫移	書 到完部／書到日如律令／卒史山書		146A
丞萬年謂大煎都候寫移書到完部／	書 到日如律令／卒史山書		164
書到完部／書到日如律令／卒史山書	下部是官承	書 從事下當用者	203
厥胡守士吏／事	下當用者如詔書		203
見決事與霸德安漢不所坐不同即上	對具		205
布地丞賢下守候城部都尉臨部官承	書 從事下當用者如詔書／到言		207
都尉臨部官承書從事下當用者如詔書	／到言		213
出 關人畜車馬器物如關	書 移官會正月三日毋忽如律令		217A
丙寅大煎都守候丞	士吏異承書 從事／下當用如詔書／	令史尊	221
士吏異承書從事／下當用如詔書	／令史尊		247+
未得是道里遠辟回往來希官簿身賤	書 佐蓬卿 用 一斤直十		247+
厚恩政叩頭叩頭幸甚 前政數奏	書 不通叩 頭 因／同吏郎今置為敦		301
家	書 為金城太守□□□□□願幼卿		335
入西蒲	書 二封 其一封文德大尹章詣大使五		336
出粟一斗二升以食使莎車續相如	書 上良家子二人癸卯口		344

(図3)「書」に関する釈文の KWIC 整列化

4. 索引の作成と漢字

漢簡の内容に関しては内容による大きな分類がいくつかできており、それらの検索とともに、キーワードによる検索が可能であり、索引が出力されている。但し、一字索引検索や基本検索によるディスプレイ出力以外は、数量が膨大になるため作っていない。漢簡に書かれている文字はすべて中国古代漢字であり、ここでは研究者が許容する範囲で JIS 漢字を用いており、全体で第一水準 14452 文字、第二水準 1695 文字、その範囲を越えるものとして 110 文字の外字を作成し、1215 文字用いている。

5. 漢簡の相方探しと冊書の復元

この様な簡牘の画像データベースからどのような研究方向があるか少し述べる。

写真本の配列を先ず、出土番号順に並び替えて、書かれている内容による分類を用いて大きく分類する。次に敦煌漢簡 951 本のうち、完全なものは 198 本しかなく、欠落した簡を選び出してちぎれた相方を探すことが考えられる。欠落した漢簡は一文字による意味の反転を防ぐため、切れ口が完全に一致するもののみ探すので、これは画像的なマッチングを考えても可能であるものと考えられる(写真本に掲載されている木簡は原寸大である)(図4)。



(図4) 漢簡の切れ口

次に冊書は同種類の本を用いて、一人の作者(当時の官吏)が書いていたものと推定できるので、書かれている共通の文字に着目して、筆跡の共通性を見いだすことにより、冊書を構成していた簡を捜し出すことが考えられる。次の例は冊書と考えられる詔書の一部に現れる「書」と「令」を原画像に処理を施して切り出したものである(図5)。



(図5) 敦煌漢簡の「書」と「令」

漢簡の釈文から推論して冊書の一部の簡と考えられる簡と他の簡から文字を切り出してみた。「書」の140, 142と「令」の143 が同一作者による文字と考えられるが、画像的特徴から見ても類似性が確かめられる⁽⁵⁾。簡牘研究において重要な木簡字典は敦煌、居延、武威、などの代表的な簡牘に現れる文字に対して人力により一文字ずつトレースし、収録しているが、その内容に疑問を持つ研究者もあり、このデータベースを用いて木簡字典の誤りを探すこともできる。

6. おわりに

中国古代漢字を扱う為にコードを体系付け、フォントを作成した上、各字に属性テーブルをつけ、異体字の処理や新字源コードとの対応を考慮した京都大学大型計算機センターの大型計算機用の康熙コードと三角編号法及び四角號碼法による入力法が知られているが、本システムではパソコンの能力の限界で用いていない。

本研究は文部省科学研究費(63450050)の補助を受けて実施したものである。

参考文献

- (1) D. Chavannes: Documents Chinois de Couverts par Aurel Stein, Oxford (1913)
- (2) H. Maspero: Les Documents Chinois de la Troisième expedition de sir Aurel Stein en Asie Centrale, British museum (1953)
- (3) 奈良国立文化財研究所における情報処理の現状, 奈文研資料(1988)
- (4) 島崎他: 漢字の字形分析と漢字入力システムへの応用 情報処理学会第26回全国大会, pp. 1153-1154 (1983)
- (5) 園田他: 木簡画像の特徴の抽出, 情報処理学会第 40 回全国大会予稿集(1989)