

科学研究費助成事業 研究成果報告書

令和 6 年 6 月 14 日現在

機関番号：34315

研究種目：基盤研究(C)（一般）

研究期間：2018～2023

課題番号：18K12001

研究課題名（和文）情報化時代における佚文収集の手法についての研究：大蔵経からの抽出を事例として

研究課題名（英文）A Study on Methods of Collecting Oraganaic Texts in the Information Age: A Case Study of Extracting Texts from the Daizo-kyo Sutra

研究代表者

山田 崇仁（Yamada, Takahito）

立命館大学・文学部・非常勤講師

研究者番号：20425010

交付決定額（研究期間全体）：（直接経費） 2,500,000円

研究成果の概要（和文）：本研究は、漢字文献のテキストデータベースを利用して、現在刊本が伝わっていない佚書の文章（佚文）を効率よく収集するための方法論構築を主たる目的とした。本研究では「SAT大正新脩大蔵経テキストデータベース（以下SATと略す）」に収録される、漢字で書かれた仏教文献（以下「漢字仏典」と称す）に引用される、随以前の外典漢籍（以下「伝世文献」と称す）佚文を収集の対象として行われた。本研究では、書籍名や佚文特有のフレーズから、大規模テキストデータベースを利用した佚文収集を効率的に収集する手法を試み、従来よりも短期間に多くの佚文を獲得することができた。

研究成果の学術的意義や社会的意義

これまで手作業で行われてきた佚文収集を、大規模テキストデータベースを利用することで効率的に行うための方法論構築を研究テーマとして設定した。申請段階では「書名」「パター的なキーワードの設定」を方法論の中心部分として挙げていたが、それによる佚文収集には一定の成果をあげることができた。一方、収集した佚文の整理と評価とに問題（と言うよりも研究検討課題）が生じたが、それは通常の文献検討と同じ校訂に類する部分であると評価した。本研究成果として作成している佚文集では、参考情報明示的な性格も兼ね、多少疑わしいものも（それを指摘しつつ）収録し、個々の佚文の最終的な評価は研究者個人に委ねる方針を立てた。

研究成果の概要（英文）：The main objective of this study was to construct a methodology for efficiently collecting lost texts (佚文) that have not been handed down to the public yet, using a text database of Chinese character literature. This study was conducted using the "SAT Taisho Shisyu Daizokyo Text Database" (hereinafter referred to as "SAT"), which is a database of Buddhist texts written in Chinese characters (hereinafter referred to as "Buddhist texts written in Chinese characters").

In this study, we attempted an efficient method of collecting lost texts using a large-scale text database based on book titles and phrases unique to lost texts, and were able to acquire a large number of lost texts in a shorter period of time than in the past.

研究分野：中国先秦史

キーワード：漢字 佚文収集 出土文字資料 大蔵経 先秦史資料 情報検索方法 情報検索方法 輯佚書作成

1. 研究開始当初の背景

中国古典学の諸分野において、現在失われた文献の一部を探求する佚文収集作業とその成果たる輯佚書の刊行が行われ、斯界を大いに裨益してきた。これらの書物からの佚文収集は、漢唐期の注釈や大規模類書を対象に手作業で行われたものであるため、効率性に佚文を収集することが難しかった。

これに対し研究代表者は、引用文献名が明示されていない佚文を収集するために、情報機器を利用したテキストのパターン分析を行い、それを利用した佚文収集を行った。その結果、系譜史料などのパターン化された記述が多い佚文収集には、その手法が有効であることを明らかにした。また別に漢字仏典に含まれる伝世文献の記述を利用した研究を行ったが、その調査過程で佚文の存在に気がつき、「SAT大正新脩大藏經テキストデータベース」(<https://21dzk.l.u-tokyo.ac.jp/SAT/>)を利用することで効率よく佚文が収集できるのではないかとする考えに至った。それを実際に試した研究では、「佚書の書名・著者名」や「既存の佚文」をキーワードとすることで、多数の佚文収集に成功している。

しかし、佚文収集の過程で、上記の方法では見落としていた佚文をいくつか見出すことになった。そのため、特定の書物群を対象に悉皆調査をし、先の研究で利用した手法で収集出来なかった佚文がどのくらいの割合で存在するのかを調査し、従来の手法がどの程度有効なのか、またその方法で漏れた佚文を収集するにはどのようなキーワード設定が必要なのかを検証したいと考えようになった。それが本研究の根幹をなす問いである。

2. 研究の目的

本研究の目的は、(1)従来の方法による佚文収集・(2)特定書物との比較による方法論の検証・(3)(1)(2)を踏まえた新たな方法論の提示の三点であり、加えて、独自性と創造性として、佚文収集の方法論と収集対象文献の二点を設定した。

近年の佚文収集手法として、大規模テキストデータベースの利用が珍しくないが、検索キーワードとして佚書の書名や著者名が使用されることが一般的である。それ自体は佚文の正当性担保のためにも当然のことだが、それらが明示されていないあるいは異なる名称で引用される場合もあるため、検索漏れがどうしても生じてしまう。従来は、その問題を文献の悉皆調査で解消していたが、本研究では悉皆調査の結果から双方方法で収集した佚文を比較し、その結果から検索漏れを出来るだけ少なくするキーワード設定について調査・検証するという方法論の確立こそが独自性となる。

3. 研究の方法

◆ (1) 佚文収集

既に研究代表者は、大規模テキストデータベースを利用した伝世文献の分析による佚文収集について一定の研究成果を挙げている。本研究では、それを踏まえたより大規模テキストデータベースを利用したより効率的な佚文収集の手法を確立し、実際に漢字仏典からの佚文収集に取り組んだ。

◆ (2) 悉皆調査にもとづく佚文の収集と、キーワード検索との比較・検証

本研究では、書名・著者名をキーワードとして検索を行い、多くの佚文を収集する一方、収集の過程でそこから漏れた複数の佚文の存在に関して、手作業で悉皆調査して収集し佚文とを比較し、キーワードによる検索を利用した佚文収集がどのくらい有効であるかの検証を行う。

その成果を反映することで、漏れを少なくした佚文収集作業を実施する。

◆ (3) 南宋大藏經・日本古写經と対照した異文・佚文の収集と既存の佚文との校勘

嘗ての佚文収集作業において、少なからず誤植と思われる記述を見つけた。このような誤記は、検索キーワード選定に影響を与える可能性が想定される。そのため本研究では、SATと大正蔵の対校に加えて、我が国の古写本や南宋の開寶蔵～高麗蔵～大正蔵という版本の系譜に属さないテキストと比較・校勘する。

その作業の中から、テキストの違い(異文・誤字脱字)から見出された有効だと想定されるキーワードを選定し、それをフィードバックして佚文収集を充実させる。

4. 研究成果

まず、本研究は当初3年間を予定していたが、2年目の末以降、新型コロナウイルスの猖獗により、様々な部分で研究に支障が生じた。特に特定の研究機関に赴いての資料収集やデータベース利用などに制限がかかり、新たなデータの収集に問題が生ずる事態となった。それに加えて、研究代表者が新型コロナウイルスに罹患する事態に陥ったため、一定期間研究に支障が出てしまった。そのため、度重なる研究機関の延長を申請し、それによって研究環境改善を期待し、延長

された期間で既存データの整理を行うことにした。

研究環境上の制限については、それまでに収集した情報を精査する時間が取れたと切り替え、(結果として収集した佚文のテキストにはそれほど大きな影響は与えなかったものの)中国の古代史研究、特に六朝期のそれを含む佚文や、結果としてそれ以降の先秦史の歴史観に一定の影響を与えたことを確認できたのは幸いであった。研究代表者の新型コロナウイルス罹患については、それが研究最終年度であったことや軽傷で回復したこともあり、回復後に研究を再開することができたため、幸いにして致命的な影響は出なかった。

◆ (1). (2). 佚文収集と比較検証

研究計画に基づき、成果を簡単に整理しておく。

まず研究の調査部分として、既存研究である程度の有効性を示した方法論(書名・既存佚文から見出したキーワードなど)を利用し、大規模データベースでの検索を行った。

無論、単純に検索するのみでは、大量の不要なデータ(所謂ゴミデータ)が集まるだけであり、そこからどのようにして情報として有為な(佚文である可能性が高い)テキスト群を抽出するかが問題であった。

本研究では、PCDAサイクルの方法論を用い、適宜方法論のアップデートとキーワードの修正を行いつつ、より確度の高い佚文を収集する試みを行った。

その中で得た知見を箇条書きに整理したものが以下の三項目である。

データベースの質に起因する問題

既存の佚文集との比較

収集したデータは本質的に何を表現したものなのか

まず から述べる。これは、『四庫全書』や『四部叢刊』と言った大規模叢書を対象とした大規模データベースが一般にも利用可能になったことや、中国哲学書電子化計画(<https://ctext.org/zh>)などのテキスト量的には相当に充実してきた一方、元のテキストデータの出自が不明+それほど質が良くないという問題点である。

本研究の主要なデータ収集対象であるSATは、底本として業界標準の『大正新脩大藏經』を採用し、データの校正をしっかりと行い、更にはUnicode未収録字への対応も充実しており(むしろ未収録字をUnicodeに積極的に申請登録する活動を行ってきた)、加えて『大正新脩大藏經』や他の諸版との比較可能なシステムが研究期間内に構築されており、これに限っては問題はそれほど無かった(世界的に見て、むしろSATが先進的かつ学術向けに必要な部分を実装していると評価すべきだろう)。

問題は、佚文をチェックするために利用した他のデータベースである。まず、底本情報の欠如やテキストの校訂が不十分なものが多いのは当たり前で(『四庫全書』や『四部叢刊』のような数百万字以上の大規模テキスト群を扱っている、あるいは有志による入力やOCRでの文字認識に頼っている以上、校正がややおろそかになるのは仕方が無い部分ではある)、単純に検索 テキスト収集では作業が終わらず、底本や校訂が行われている標点本の類と比較する作業が必須となった点である。

この点については、とりあえず検索で集めるだけ集め、底本や筋の良いと判断した標点本や校訂情報を集めた書籍と比較検討し、それを収集した佚文データと比較するスタイルを採用することになった。この辺りは、研究の中でこれまでも行ってきたテキスト校訂作業とほぼ同じ過程であり、手間がかかる一方、電子テキストはそのようなものだと割り切って研究遂行している部分であった。

次に である。これは特に、緯書の佚文収集整理作業で直面した。緯書は、現存する書籍が少なく、佚文を収集して研究が行われている。その一大成果であり標準的な輯本といえるのが『重修緯書集成』である。本研究でも、SATから収集した緯書の佚文を『重修緯書集成』と比較する作業を行ったが、『重修緯書集成』と書名が異なり内容が一致するものが少なからず存在した。これは単純に書名間違いではなく、同じテキストを複数の緯書が使い回していたことを示す可能性があると考えている。この部分の整理については、報告書執筆段階でも定案が決まらず、後日整理したものを公開予定としている。

最後に である。こちらは、佚文それ自体の性格に関連する問題である。それは、書名を冠する佚文は、そもそも本当にテキストからそのまま抜き書きしたもの否か、という佚文収集の本質的な問いに関わっている。これは、佚文とされる部分を引用した人物が、自らの執筆スタイルに沿って引用を改編した可能性があるためである。

特に研究代表者が佚文抽出のキーとして設定した「パター的な内容」について、それが佚文そのものではなく、引用者が整形したものである可能性に理解が及ぶようになった為である。その結果、佚文そのものではないが、そのエッセンスが含まれるテキストとして、評価すべきテキストの評価について考えるようになった。

◆ (3). 南宋大藏經・日本古写經と対照した異文・佚文の収集と既存佚文との校勘

SATに拠って抽出した佚文を、他の版本や写本と比較し、より古いテキストの様相を明らかにしたいというのがこの項目のテーマであった。

しかし、筆者が比較対象とした佚文の文章は、偶然の可能性もあるが、著しい違いを見せたも

のは無かった。これは、筆者がたまたま佚文収集の対象とした仏典諸篇が、大正蔵の底本である高麗蔵が底本とした最古の木版印刷大蔵経である北宋：開宝蔵と同系統（開宝蔵）の写本によって日本古写経が抄写されてきたことによると考えている。そのため、筆者が比較対象とした部分は、少なくとも北宋初期には定着したものと結論付けた。

また別に、歴代三宝記に見える東周王の諡・諱について、一部『史記』と異なる表記が見られた事に着目し、分析を行った。その結果、それらは何れも六朝期の『史記』古写本や注釈に系統に由来するものであり、『史記』より古い資料の引用と言うよりも、六朝期の『史記』テキストの表記を伺わせる資料として扱うべきとの結論を得た。

更に、一部の史伝部に見える（『史記』には存在しない）夏殷周の君主在位年情報もまた、六朝の上古史研究を反映したものが仏典（中国史と仏教の歴史の年代を記述する際に、六朝期の上古史が提示した独自の在位年情報）を通じて残され、前近代の日本の歴史観にも影響を与えたことを明らかにした。

これらの研究の結果から、仏典に引用される佚文を含む外典漢籍の記述に見える固有名詞などの異表記は、先秦期のものであることはほぼ無く、南北朝時代のテキスト群の情報を保持する可能性があり（既存の外典漢籍の校勘の流れから切り離され存置されてきた）、それは場合によっては既存の漢籍のテキストを校勘する情報源となり得る可能性を見出した。

一方、それらの部分については、筆者・板刻字の誤字故に発生したとおぼしき異表記も見られることから、その評価には慎重な検討が必要となることを指摘しておく。

まとめ

本研究での中心課題である「大規模テキストデータベースを利用した効率的な佚文の収集」は、「研究成果」の項目で述べたように、データベースに起因するテキストの質という問題はあるものの、佚文収集作業自体には一定の成果が得られた（既存の佚文集収録済みのものはもとより、未収録の／言及が余りされないものもいくつか発見することができた）。

基本的に、書名やパターン記述をキーワードとして検索する手法は有効であることが確認できた。また、その部分の周辺を注意深く読み解くことで、直接の引用が明示されていないものの、佚文と関係が高いテキストを複数獲得することが出来た。厳密に言えばそれらは佚文と認めるには問題があるだろうが、この点について研究代表者は、より多くの情報を提示することを重視し、それを見た他の研究者が自らの判断に於いて取捨すれば良いと考えている。

そこで獲得された先秦期について述べる文献佚文の多くは、当初期待したような先秦期（或いは劉向校書以前）の資料に由来するものではなかった。それらの多くは後漢末～三国・西晋～六朝期に盛んに行われた『史記』記述の克服を目指して行われてきた著作群の引用であり、その点では申請当初に期待した先秦期文献の佚文収集という目的は達成できなかったという結論になる。

しかし、それら後漢～六朝期の先秦史研究文献の佚文は、『史記』などの現存する伝世文献のより古いテキストの状態を伺うための材料となり得るため、先秦史研究にとっても貴重な研究資料となることは間違いない。更にはこの時期に始めて登場する（史記が記述しない）夏殷周君主在位年情報という興味深い内容（それ自体の典拠は不明／むしろ想像の産物である可能性が高い）について、後世に与えた影響など先秦史に対する歴史認識関連の研究材料として、本研究とは直接的な関連はないが、先秦史や史学史研究に裨益する部分が少なくないと考えている。

5. 主な発表論文等

〔雑誌論文〕 計11件（うち査読付論文 6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 山田崇仁	4. 巻 11
2. 論文標題 屈子赤目ほ	5. 発行年 2023年
3. 雑誌名 漢字学研究	6. 最初と最後の頁 81-95
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 -

1. 著者名 山田崇仁	4. 巻 10
2. 論文標題 晉公盤	5. 発行年 2022年
3. 雑誌名 漢字学研究	6. 最初と最後の頁 85-109
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 -

1. 著者名 山田崇仁	4. 巻 10
2. 論文標題 講演 中国古文字のデジタルテキスト化に関する諸問題	5. 発行年 2022年
3. 雑誌名 漢字学研究	6. 最初と最後の頁 25-46
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 山田崇仁	4. 巻 9
2. 論文標題 呉王餘昧劍	5. 発行年 2021年
3. 雑誌名 漢字学研究	6. 最初と最後の頁 61-81
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 -

1. 著者名 Takahito YAMADA	4. 巻 4-2
2. 論文標題 The History of the IT Environment of Sinology in Japan	5. 発行年 2021年
3. 雑誌名 The International Journal of Chinese Character Studies	6. 最初と最後の頁 111-129
掲載論文のDOI (デジタルオブジェクト識別子) 10.18369/WACCS.2021.42.111	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 山田崇仁	4. 巻 8
2. 論文標題 金文通解 王子午鼎	5. 発行年 2020年
3. 雑誌名 漢字学研究	6. 最初と最後の頁 21-35
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 -

1. 著者名 山田崇仁	4. 巻 8
2. 論文標題 古文字学研究文献提要“陳劔の論考より”：拋郭店簡積読西周金文一例	5. 発行年 2020年
3. 雑誌名 漢字学研究	6. 最初と最後の頁 161-165
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 山田 崇仁	4. 巻 7
2. 論文標題 趙平安著「“達”字兩系説 兼釋甲骨文所謂“途”和齊金文中所謂“造”字」 [附] “達”字“針”義的文字學解釋 從一個實例看古文字字形對詞義訓詁研究的特殊重要性」	5. 発行年 2019年
3. 雑誌名 漢字学研究	6. 最初と最後の頁 14-15
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 山田 崇仁	4. 巻 3
2. 論文標題 聖武天皇第一皇子の諱について 避諱研究の一環として：「基」か「某」か	5. 発行年 2020年
3. 雑誌名 日本漢字学研究	6. 最初と最後の頁 -
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 -

1. 著者名 山田崇仁	4. 巻 51
2. 論文標題 『世本』作篇考	5. 発行年 2019年
3. 雑誌名 花園大学文学部研究紀要	6. 最初と最後の頁 横41-66
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 山田崇仁	4. 巻 10
2. 論文標題 『國語』章昭注引『世本』系譜集成	5. 発行年 2019年
3. 雑誌名 中国古代史論叢	6. 最初と最後の頁 印刷中
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4件(うち招待講演 2件/うち国際学会 2件)

1. 発表者名 山田崇仁
2. 発表標題 關於日本中國學的IT利用環境的歷史(日本中國學におけるIT利用環境の歷史について)
3. 学会等名 世界漢字學會第八屆年會(招待講演)(国際学会)
4. 発表年 2021年

1. 発表者名 山田崇仁
2. 発表標題 中国古文字のデジタルテキスト化に関する諸問題
3. 学会等名 漢字文化の展望（白川静記念東洋文字文化研究所：オンラインシンポジウム）（招待講演）
4. 発表年 2022年

1. 発表者名 山田崇仁
2. 発表標題 『歴代三寶紀』帝年に記される周王の諡號・諱について 佛教史書に記される先秦史関連資料研究の一環として
3. 学会等名 東アジア文化交渉学会（国際学会）
4. 発表年 2020年

1. 発表者名 山田崇仁
2. 発表標題 聖武天皇第一皇子の諱について
3. 学会等名 日本漢字学会
4. 発表年 2018年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

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

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1件

国際研究集会 中国史史料研究会 第二回オンラインセミナー「中国史学から見る「中国」概念」	開催年 2023年～2023年
---	--------------------

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
---------	---------