

令和の大蔵経と次世代人文学

本研究は、研究媒体のデジタル化（DX）と人工知能（AI）が共同して急速に創成しつつあるあらたな研究環境の整備において、日本の人文学が対応すべき要件を摘示するとともに、それらを実現するデジタル学術空間のモデルを構築し、人文学の伝統的営為を発展的に再編するための具体的道筋をしめす――

公式サイト



2026.2
Collaborative
Research Data
Infrastructure
for the Humanities

第4号

<https://creativecommons.org/licenses/by-sa/4.0/deed.ja>

<https://reiwadzkhii.asia/>

1 「大蔵経の編纂」についての個人的感傷

●木越 康 [大谷大学]

おそらく多くの年配の仏教学者がそうであるように、大蔵経関連の仕事にはじめて関わったのは「索引づくり」と呼ばれた作業への協力でした。当時大学院生であった私がようやく「大正新脩大蔵経」と格闘しての本格的研究をはじめた頃です。カード一枚に手書きで文字を書き写して情報を記入するという、途方もない作業でした。次に関わったのがSAT データベースのチェック作業です。すでに大学教員でしたので、学内の大学院生を集めて作業を管理する仕事を担いました。今回の「令和の大蔵経の編纂」事業は、したがって私にとっては三度目の仕事となります。還暦を越えた私にはおそらく最後の関わりであり、大蔵経へのご恩返しのお機会となることを願っています。

2025年6月に開催されたキックオフ・シンポジウムでは（本誌第1号参照）、集まった仏教学や情報学の専門の先生方から様々な質問と意見が寄せられました。研究代表者・分担者の方々は一つひとつに丁寧にお答えになり、より良い研究の方向性を共に見出そうとしてくださいました。その時私が抱いた疑問は「そもそも令和の大蔵経とは何であるのか」というものです。新しい技術によって獲得される膨大な画像データやテキストデータが最終的にどのような形で集約されるのか。「令和の大蔵経」とは、何がしかの形で全体を集約した「真正データ」の成立をもって完成とするのか。あるいは完成に向けて収集されるさまざまな様式の膨大なデータが「それ」であるのか。この問いは、実は今も明快な回答を得たわけはありません。しかし今は、この問いは解けないままにいる方がよいのだとも考えるようになりました。それは「令和の大蔵経を求めろ」という活動自体が、「釈尊とは誰か」「釈尊の教えはどこにあるのか」を求めた仏弟子たちの歴史への参画という意味を持ち得ると考えるからです。

2025年11月末には、京都の大谷大学で「第1回令和の大蔵経編纂作業者会議」が開催されました（本誌第3号参照）。東京から主要メンバーも来学し、佛教大学と龍谷大学、大谷大学のスタッフ総勢50名ほどが集まり、作業の確認を行いました。大学院生たちにはここではじめて「令和の大蔵経編纂 RA（リサーチ・アシスタント）」という名称が与えられました。実は大谷大学では、プロジェクトがはじまった頃から「これは令和の仏典結集か？」という声が挙がりました。抑揄なのか称讃なのかわかりませんが、恐れ多いことだと思います。しかし今では、懸命に取り組む学生たちを見て、あながち間違いでもないと思うようになりました。すべての作業が釈尊の教えをを求める行為であるとするならば、それは仏典結集の精神を汲むものとして認識すべきでしょう。

責任ある諸学者による研究であり、大切な資金を頂いての活動であるので、当然、成果としての「令和の大蔵経」は完成させなければなりません。しかし個人的には、その完成は仏教経典への敬意を込めて、ある意味カッコ付きの完成形であることを自覚しなければならぬとも思っています。それは、完成された大蔵経は〈不確かなもの〉であり、また〈間違い得るもの〉であるということです。先にこの事業への関与が私の大蔵経への最後の恩返しであるとも書きましたが、「恩返し」と言いたい一番の理由は、釈尊のお仕事の邪魔をしてはならないという思いを持つからです。それがまた、やがて出される私たちの成果に対して、予め〈不確か〉や〈間違い得る〉と述べておきたい理由でもあります。この仕事は、過去の遺産としてある仏典を、単に未来へ遺すためのものではないでしょう。釈尊の言葉に人間の生きる真実を求め続けた仏弟子たちの歴史への参画です。古い時代を知る仏教者の戯言として、聞き流していただきたい。

2 仏教研究とデジタル・ヒューマニティーズ —国際シンポジウム「DX-AI 時代における人文学の意義と研究基盤の構築」開催報告

● 朴賢珍・永崎研宣

本特別推進研究の主催により、人文学 DX の未来を展望する国際シンポジウムを 2025 年 12 月 20、21 日の 2 日間で開催した (<https://sites.google.com/view/buddh2025/>)。国内外の専門家が一堂に会し、AI 技術と文献学の融合、持続可能な研究基盤の設計と運用について報告・議論が行われた。前号の全体報告に続き、2 日目午後のセッション「仏教研究とデジタル・ヒューマニティーズ」について詳報する。

仏教研究におけるデジタル化の第一線で活躍する 6 名が登壇し、各プロジェクトの知見と成果、ならびに課題と可能性が多角的に示された。多様なアプローチが提示された一方で、研究の信頼性を支える標準化と、長期運用を見据えた設計・協働の必要性が共通の論点として浮かび上がった。

■ 中世中国文献データベース・プロジェクトの紹介：現行の機能と今後の展望

Prof. Christoph Anderl (Ghent University, Belgium)

アンデル教授は、10 年以上にわたって進めてきた「中世中国文献データベース (Database of Medieval Chinese Texts)」を紹介した。本プロジェクトは敦煌写本を中心とした中世漢文資料、特に 9～10 世紀の口語的要素を含む非正典文献を主な対象とし、再校訂を経た高品質なデジタル版の構築と公開を行っている。TEI/XML マークアップ標準を適用し、現在までに約 250～300 点の写本をデジタル化している。

当初はマークアップ作業の副産物であった異体字コレクションが、現在では本プロジェクトの主要な柱へと発展した点も強調された。対象範囲を 4～9 世紀の年紀写本や敦煌の字書、さらにはハイデルベルク・アカデミーの仏教摩崖碑文プロジェクトから提供された約 6,000 字の異体字へと拡張することで、字形の変遷を立体的に比較研究することが可能になった。

今後の展望として、ユーザーが未知の異体字画像をアップロードすると、蓄積されたデータと照合して候補を提示する AI 支援モジュールの導入など、さらなる機能拡張の計画が示された。

■ AI による仏教知識グラフの構築：辞書的概念から典籍関係トリプルへ

Prof. 王昱鈞 (台湾・法鼓文理學院)

王教授は、膨大な仏教概念を構造化・可視化するための「仏教知識グラフ」構築に関する革新的なワークフローを紹介した。

本プロジェクトは、まず陳義孝編『佛學常用詞彙』、丁福保編『佛學大辭典』、『佛光大辭典』、および台湾教育部

『國語辭典』の 4 つの主要辞典をソースとし、LLM (大規模言語モデル) を活用して重複する語義を統合。約 5 万件のエンティティ (実体: 一意に識別される人物・場所・概念などの単位) と 6 万件の語義を整理した。特筆すべきは、単なる自動化に留まらず、AI による類似度スコアが閾値を下回った場合に人間が介入する「Human-in-the-loop」方式を採用し、学術的な厳密さを担保している点である。

構築プロセスは、①語義統合、②要約、③分類、④関係抽出、⑤トリプル (主語-述語-目的語) 生成の 5 段階で構成される。特に「分類」工程では、一般的な固有名詞 (人名・地名など) よりも、仏教研究において重要度の高い「教理」に焦点を当て、教理、人物、場所、儀礼、文献、事象などを含む 13 のカテゴリを定義した。この分類作業には、3 つの AI モデル (ChatGPT/Claude/Gemini) を併用するマルチモデル戦略を採用。各モデルの回答に基づく多数決投票方式を導入することで、単一モデルのバイアスを排除し、人間との一致率約 95% という極めて高い信頼性を実現している。

さらに王教授は、従来の RAG (検索拡張生成: 回答時に外部情報を検索して利用することで誤回答を抑える手法) が抱える概念間の関係性把握の限界を指摘し、知識グラフを構造的コンテキストとして組み合わせる「Graph RAG」の有用性を強調した。

■ キーワードから意味論へ: CBETA コーパス検索機能開発の経験

Prof. 洪振洲 (台湾・法鼓文理學院)

本発表では、CBETA コーパスにおける検索機能強化の歩みと知見が報告された。CBETA では月間平均 37 万件以上の検索が行われており、極めて高い利用需要がある。一方で、仏教文献の豊饒さと複雑さ、そして長い歴史の変遷ゆえに、従来のキーワード検索には限界があるとし、課題を「キーワード選定の困難さ」「概念表現の多様性」「異体字」「検索結果が多すぎることによる探索の困難」の 4 点に整理した。

その解決策として、まず (1) KWIC 機能、(2) 時代・著者・経録などで結果を集計・絞り込むファセット検索、(3) 異体字辞書との連携による検索漏れの防止、(4) 類義語拡張 (近似表現の抽出) などを順次導入することで、単純なキーワード一致を超えた探索を段階的に実現してきたと述べた。

さらに、キーワード検索の限界を克服するための試みとして、RAG を用いた「セマンティック検索 (ベクトル検索)」の実装事例を提示した (2024 年 5 月版公開)。これは自然言語での問いを入力すると、段落単位でベクトル化したテキストから意味的に近い箇所を抽出し、回答生成に活用する仕組みである。

■冬眠し、そして再び目覚めるには：オープンエンド型編集プロジェクトからの教訓

Dr. Patrick McAllister (Austrian Academy of Sciences, Austria)

マカリスト博士は、資金提供の終了やチームの交代により活動が停滞しがちな「オープンエンド型」プロジェクトの生存戦略について、自身が携わる対照的な2つのプロジェクトを事例に論じた。

1つ目は、南アジア研究文献の電子テキスト化プロジェクト「SARIT」である。データ (TEI XML) 自体は健在だが、ソフトウェアの依存関係の問題でフロントエンドが機能不全となり、更新不能な「冬眠」状態に陥った教訓が共有された。

2つ目は、20年以上続く「Jinendrabuddhi」の校訂プロジェクトである。現在は、単なる印刷物の刊行を超え、全体を統一されたデジタル校訂本として再構成すべく、第5章から TEI XML を直接導入し、データの統合と高度化を推進している。

博士は、プロジェクトの停滞を「死」ではなく再始動を待つ「冬眠」と定義し、特定のソフトに依存しない「VGEST (Vienna Encoding Guidelines for Editing Sanskrit Texts)」の構築を解決策として提示した。ソフトウェアはいずれ作り直す必要があるため、「データそのものの品質」と「交換可能なツール設計」こそが、AI 時代における DH プロジェクトの持続可能性を担保する要諦であると締めくくった。

■パーリ語本文批判のためのデジタルツール：四つの貝葉写本伝承に基づくパーリ仏典の校訂版作成

Dr. Bunchird CHAOWARITHREONGLITH (Dhammachai Tipitaka Project, DCI Center for Buddhist Studies, Thailand)

本発表では、4系統 (シンハラ、ビルマ、クメール、タイ等) のパーリ語貝葉写本を一次資料とする新たな校訂版を作成するワークフローが報告された。プロジェクトでは、総計300万頁規模に及ぶ写本画像のアーカイブを基盤に、専用システム「ODEM」を用いた厳格な転写工程を構築している。具体的には、2人の作業員による二重入力と自動比較、さらに第三者によるチェックを組み合わせ、精度を極限まで高めている。

校訂作業では、語単位のテキスト整列を行う「ITAP」と独自開発の編集プログラムを活用し、本文確定と校異作成を効率化している。現在は冊子体での刊行を優先しているが、今後は同期の問題を解決するため、オンライン公開を先行させる運用計画も示された。

また、今後は AI 技術の導入を検討しつつも、巨大なコーパスでは自動処理の精度が99%に達したとしても膨大な誤りが残るといった課題を指摘した。そのため、最終的な品質管理には人手による検証が不可欠であるとし、技術と専門家の知見を統合した逐次公開を推進する方針を強調した。

■翻訳・OCR・意味検索：Dharmamitra エコシステムの現状と今後の展望

Prof. Sebastian Nehrlich (東北大学)

ネルディッチ教授は、AI 技術が仏教学に与える影響を「タスクのピラミッド」として整理し、現状と展望を報告した。ピラミッドの下層にある OCR やレイアウト解析といった成果の定義が容易な領域は自動化が進む一方で、統計的な確率に基づく現在の AI では、「より難しい読法」を選択するような厳密な本文批判 (校訂作業) の自動化は依然として困難であるとの見通しを示した。

その上で、翻訳・検索・OCR を統合した研究プラットフォーム「Dharmamitra」を紹介した。2025年2月より Gemini API を活用したシステムへ移行し、パーリ語、サンスクリット語、チベット語、漢文を横断する多言語機械翻訳を実現している。機能面では、文法解析や辞書リンクを提示する「英語 (説明付き)」や、4言語の並行箇所や注釈書を一括で提示する「英語 (リサーチ)」など、研究者のワークフローを強力に支援するツール群をデモと共に提示した。

また、生成 AI 特有の「ハルシネーション (幻覚)」については、意味検索技術 (セマンティック検索) との組み合わせにより抑制を図りつつも、最終的にはユーザー自身による原典照合と検証が不可欠であるとする設計思想を強調した。

最後に、BDRC との提携による大規模データの統合や、次世代の探索システム「Metra Explore」の開発計画など、2026年以降のロードマップを示して発表を締めくくった。

■総合討論：持続可能なデジタル学術基盤をめざして

最終セッションでは、「持続可能性」を軸に、AI 時代の研究基盤のあり方について活発な議論が交わされた。

とりわけ TEI をめぐっては、導入や入力コストの高さが長年の課題である一方で、学術校訂版としての分析枠組みを保持し、データの長期的な互換性を支える「共有標準」としての不可欠な意義が再確認された。その上で、作業の効率化を図るべく、扱いやすい内部形式から TEI へ変換するコンバータの構築や、Cursor 等の AI 支援型エディタを活用してマークアップの負担を軽減するといった現実的な方策が共有された。

AI による意味検索 (セマンティック検索) については、ユーザーの行動が「キーワードがどこにあるかを探すこと」から「直接的な答え (知識) を得ること」へと変化している点が指摘された。特に、多くのユーザーが詳細な出典リストを確認するよりも、AI が生成した「短く要約された回答」をそのまま信じてしまいがちな現状に対し、システム側が根拠を明示する「説明責任」や、利用者が AI の限界を正しく理解するためのリテラシー教育の必要性が強調された。

また、仏教知識グラフの構築において、経典からの AI による直接抽出は文法的関係に留まる限界があるのに対し、専門家が編纂した「仏教辞典」をデータソースとすることで、より正確な教理的関係を記述できるという実務的な知見も共有された。

最後に、技術の進歩を柔軟に取り入れつつも、人文学的な厳密さをいかに担保し続けるかという問いを共有し、2日間にわたるシンポジウムを締めくくった。

■参考リンク一覧

●Database of Medieval Chinese Texts <https://www.database-of-medieval-chinese-texts.be/> ●CBETA <https://cbeta.org/> ●SARIT <https://sarit.indology.info/>

●Vienna Encoding Guidelines for Editing Sanskrit Texts <https://www.oewav.ac.at/en/digital/digital-projects/vegest>

●Dhammachai Tipitaka Project <https://www.dhammachaitipitaka.org/> ●BDRC <https://www.bdrc.io/> ●AI 仏教研究プラットフォーム「Dharmamitra」<https://dharmamitra.org/ja>

3 公開学術集会から——DHの最前線 ●王一凡・永崎研宣

本研究課題では、令和大蔵経編纂に関わる様々な要素について国内外で広く議論を行ない、仏教学のみならず人文学全体にも資することを目的として、仏教学及びデジタル・ヒューマニティーズに関する公開学術集会を開催してきている。また議論の場を広く公開するために、基本的にはオンライン同時配信も実施している。これらについて、以下に簡潔に報告する。

2025年11月22日(土曜)～23日(日曜) [講演] [発表]
●復旦大学日本研究センター第35回国際シンポジウム「AI時代のデジタル・ヒューマニティーズと日中学术交流(人工智能时代的数字人文与中日学术交流)」



日本と中国の研究機関より総勢24組の発表者が一堂に会した大規模な国際シンポジウムが、復旦大学日本研究センターにて開催された。

はじめに、基調講演として3名が登壇した。中国社会科学院の高洪氏は、現在のAI技術の原理と認知科学の知見を踏まえ、AIの可能性と限界、研究者とAIの協同のあり方とリテラシーについて語った。本研究課題の研究分担者・永崎研宣は、SATの取り組みを背景に、人文学の基盤であるテキスト作成を支援するためのAI利用と、人文学の知識をAIに伝えるための構造化という両輪の重要性を論じた。大阪大学の田畑智司氏は、人文学と計算機技術の長い関わりに触れながら、研究者および研究方法間の対話を促進し発展させるものとしてのDH、そしてAI活用という視点を提示した。

続いて、両日にわたる5つのセッションの中で、21件の発表が行われ、様々な事例が紹介された。本研究課題の内容または方向性に関連が深いものとしては、中国からみたSATを含む日本のDH研究の特色(王賀氏)、異体字を多く含む文献のテキスト・画像処理基盤の構築(劉冠偉氏)、RAGによる低リソース言語としての「候文」翻訳の試み(王侃良氏)、古漢字の文字情報を記述する構造化データベースの構築と可能性(片倉峻平氏)、史料による戦前日中映画界のネットワーク分析とその批判的解釈(江暉氏)、歴史上の人物を模倣したチャットシステムの設計と開発手法(李恩明・莊麦氏)などが挙げられる。また、本研究課題の研究分担者・大向一輝が東京大学におけるDH教育の方法論について、事務局からも、王一凡が統一経音義字形データベースとその設計理念について、高須賀萌が画像と文字の複合テキストである黄表紙のTEI構造化について発表した。

高度に専門的な議論が交わされたにもかかわらず、会を通じて復旦大学翻訳学科による極めて高品質な双方向通訳が提供されていたことも特筆に値する。

また、会に先立つ11月21日(金曜)午後には永崎による本研究課題の紹介を含む講演とTEIに関するワークショップが開かれた。こちらも会場は盛況で、中国の参加者から深い理解に基づく質問もあり、関心の高さが伺えた。

本シンポジウムは、本研究課題の国際的な認知を高めると

もに、中国で進行中のプロジェクトや研究者との交流を深める契機を作る得難い機会となった。

2025年12月13日(土曜)～14日(日曜) [参加]
●人文科学とコンピュータシンポジウム(じんもんこん) 2025
<https://jinmoncom.jp/sympo2025/>

情報処理学会人文科学とコンピュータ研究会が毎年主催する人文科学とコンピュータシンポジウム(じんもんこん)2025が、本年は、九州大学伊都キャンパスにおいて開催された。今回のテーマは「社会をつなぎ直す情報技術」であった。

このシンポジウムはいわゆるデジタル・ヒューマニティーズで日本を代表する査読付きシンポジウムという位置づけであり、人文学における様々な分野におけるデジタル技術の適用に関する発表が集まる。英語での発表も可能なため、海外からの発表も複数みられ、全体として60件ほどの発表があった。本特別推進研究の主要なテーマがテキスト研究であることから、ここではテキスト研究を扱った発表をおおまかに見てみたい。

今回のテキスト研究系の発表は、資料のデジタル化と注釈付け、大規模テキストの計量分析、構造化・標準化にもとづく再利用性の向上、という3つに分類できそうである。まず、翻刻や市民参加型の注釈活動を対象に、作業過程や思考の痕跡をデータとして捉え、品質管理や合意形成、参加者の行為分析へ接続しようとする流れがある。次に、現代のSNS投稿から政策形成に関する議事録、テレビニュース、戦時期の評論まで、異なる媒体・時代・ジャンルのテキストを同じ分析枠に載せ、埋め込み表現やトピックモデル、通時・共時比較によって言説の変化や主題構成を可視化する試みが目立つ。物語研究では、登場人物の台詞や評価コメントなど、作品外部の反応も含めたデータを用いて、魅力・評価・組織像の形成要因を抽出し、読者との相互作用を定量化する方向が示される。古典籍や前近代資料については、TEIなどの記述の枠組みを前提に、固有表現の関係付けや引用・発話情報の付与、編纂フローの設計といった「テキストを機械可読な形で研究できるようにする」工程そのものが研究発表の対象となっている。加えて、くずし字などの前近代の日本語の表記や字形の学習支援、欠字処理、OCR結果の構造化など、文字レベルの課題を扱うものもあった。全体としては、解釈の厚みを保ちながら、データ化・分析・基盤化を往復し、再現可能なテキスト研究へ近づきつつあるかのように思えた。

テキスト研究系以外にも、花押の画像検索、源氏香図に着目した香りの表現、古楽譜の解読、江戸時代の大規模空間データセットの構築、狩野派粉本の画題、3Dモデル作成時の課題など、人文学における様々なタイプのデータを作成したりそれを分析したりすることに関する研究発表が行なわれた。このシンポジウムではポスターセッションが用意されており、2回にわたって20件ほどのポスターの前にそれぞれの発表者が待ち受け、参加者達と様々な議論が賑やかに展開された。このシンポジウムでは仏教学からの発表がこれまで様々に行なわれてきており、今後は本特別推進研究に関連する発表が増えていくことを期待したい。

DHの最新状況をキャッチアップするためのイベント情報は「人文学情報月報」をチェックしてください

ニューズレター 令和大蔵経と次世代人文学 第4号 発行日——2026(令和8)年2月28日 編集発行人——研究代表者 下田正弘
発行——「デジタル研究基盤としての令和大蔵経の編纂」事務局 URL——<https://reiwadzkh.dhii.asia/> 刊行サイクル——隔月刊
研究課題——「デジタル研究基盤としての令和大蔵経の編纂一次世代人文学の研究基盤構築モデルの提示」 研究課題番号——25H00001