

令和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ゆる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提示するとともに、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――



1 令和大蔵経と正倉院聖語蔵本

●林寺正俊 [北海道大学]

『大正新脩大蔵経』（以下、大正蔵）は、周知の通り、底本に高麗蔵再雕版が用いられ、校本として宋版、元版、明版、旧宮内省所蔵の東禅寺版・開元禅寺版など、各種の版本が用いられていますが、一方で正倉院聖語蔵（もとは東大寺勝院の蔵経）中の奈良写経、とりわけ「天平十二年五月一日記」の願文をもつ光明皇后御願経も校本として用いられており、その異読は図1のように脚注に（聖）として示されています。

私は日本古写経の調査・研究に関わってきましたので、正倉院聖語蔵本について少しく述べさせていただきます。大正蔵の脚注に採られている正倉院聖語蔵中の光明皇后御願経とは、光明皇后が両親の追福（故人の冥福をいのること）、夫である聖武天皇の福寿、臣下の忠誠、衆生救済と仏法流布を願って作った写本一切経であり、各経巻末尾に付せられた願文の日付から一般に「五月一日経」と呼ばれているものです。底本には天平七年（735年）に唐から帰朝した玄昉の請来した経巻が主に用いられ、また天平勝宝六年（754年）頃から勘経が行われたことによってテキストとしても高く信頼され、それ以降の写本一切経の根源的存在・源流ともなった権威ある勅定一切経でもあります。最終的に6,500～7,000巻規模の一大仏書コレクションになったと推定されていますが、正倉院聖語蔵本として現存するのは750巻（当該デジタル画像は宮内庁正倉院事務所編『聖語蔵経巻天平十二年御願経』丸善雄松堂に収録）で、巻間に存するとされる約250巻とあわせても、全部で1,000巻ほどが現存するに過ぎません。こうした大変貴重な聖語蔵中の五月一日経が校本として用いられたことによって、各版本一切経の創出・流布よりも前の、八世紀に遡る写本時代の一切経の読みを知ることができるようになり、その点で聖語蔵本の読みを

示す脚注には極めて大きな意義があると言えます。

しかしながら、その上であえて申しますと、大正蔵の脚注に示される聖語蔵本の読みについては問題点もあります。一つには、大正蔵の巻、あるいは個々の經典によって、異読の採り方の精度に粗密があることです。聖語蔵の経巻すべてを確認したわけではありませんが、精度にばらつきがあることはやはり否定できません。関東大震災によって同経が東京帝室博物館から移動されたと聞きますので、そうした混乱の影響が少なからず関係しているのかもしれませんが。第二には、異読の示し方についてです。五月一日経には本文墨書の右傍にしばしば小さな朱字が記されており、これらは勘経の痕跡と推測されますが、大正蔵の脚注には聖語蔵本の異読として、墨字が採られている場合と、断りなく朱字の方が採られている場合とがあります。しかし脚注だけでは、それが墨字なのか、朱字なのかを知ることができません。第三には、特に朱字が異読として採られている場合、「墨字は誤りで、朱字が正しい」という判断のもとに採録されたものと推測されますが、しかし朱字をそもそも五月一日経の正しい読みと断定してよいのかどうかという問題があります。もしかしたら朱字は異読を示しているだけかもしれないのです。こうした朱字の性格については目下検討中ですが、現代の私たちが原稿や書類を赤ペンでチェックするのと同じような感覚でもって経巻中の朱字を扱ってよいのかどうかは、注意して見ていく必要があると考えています。

令和大蔵経の編纂では諸版本の正確なテキストデータを作成すべく多くの方々作業に関わっていますが、次の段階では、大正蔵の校本として用いられた聖語蔵本をどう扱い、どう反映させるかということも課題になると思います。

【図1】 ④魔＝摩 ⑤正＝政

2 AI を使えるものにするための 資料・知識の基盤をどう整えるか

—デジタル人文学ワークショップ「AI でつながる資料と知識」開催報告

● 大向一輝 [東京大学]

2026年2月23日(月・祝)、東京大学八重洲アカデミックコモンズにて、デジタル人文学ワークショップ「AI でつながる資料と知識」をハイブリッド形式で開催し、現地に約40名、オンラインでは最大70名程度の参加を得て終日にわたり議論を行った。本ワークショップは、デジタル人文学(DH)におけるAI活用の二つの可能性—RAG(検索拡張生成)による研究資料・知識基盤とAIの接続、および知識グラフによる一次資料・関連情報・歴史人物情報の構造化—をテーマとして、基調講演と3つのディスカッションから構成された。以下、セッションごとに概要を報告する。

■ ディスカッション1 「デジタル人文学における RAG の可能性」

午前のセッションでは、岩田直也氏(名古屋大学/国立情報学研究所)、太田葵氏(産業技術総合研究所)、佐原恭平氏(株式会社COTEN)、原田真喜子氏(都留文科大学)の4名を話題提供者として、RAG技術を人文学資料に適用するうえでの具体的な工夫と課題が論じられた。

前提として、RAGとは、大規模言語モデル(LLM)が回答を生成する際に、モデル内部の知識だけに頼らず、外部に用意した文書の集合から関連する箇所を都度検索し、その結果を参照しながら応答させる仕組みを指す。通常は検索対象となる文書を段落などの小さな断片(チャンク)に分割し、それぞれを言語処理技術を用いて何らかの意味を示すベクトル(数値列)に変換してデータベース化する。利用者の問いも同じくベクトル化され、数値的に類似した断片が取り出されてLLMに渡される、という流れである。LLMが学習していない専門資料や最新の研究成果に関する情報を扱えること、そして出典を提示できることから、人文学資料への応用に大きな関心が寄せられている。

議論の論点はおおむね3つに整理できる。第一は、人文学資料に内在する引用関係や注釈などのテキスト構造が素朴なチャンク分割で容易に失われてしまうという問題である。原文をそのまま分割するよりも、いったんLLMを用いて代名詞や前後の文脈を明示化する前処理、いわば言い換えを施すことで検索精度が大幅に向上するという実例が示され、何を処理の対象とするのか自体の検討が重要であることが確認された。第二は、専門家と非専門家では問いの投げかけ方自体の形式や長さが大きく異なるため、利用

者の把握がシステム設計の要となるという指摘である。あわせて、2,000冊規模の書籍をベクトル化した一般向け教養メディアでの運用事例も共有され、提示すべき情報をどのように階層的に組み立てるかが具体的な焦点となった。第三は、RAGを正解を導くための道具ではなく、むしろ問いを深めるための手段として活用する姿勢である。検索過程の可視化、段階的な要約、ユーザ情報の記録など、解釈の主体である人を支える仕掛けが鍵となる。RAGを対象とする従来の評価指標が解釈の多様性とコンテキストの豊かさを測れないことへの批判も共有され、RAGは万能の解ではなく、後述の知識グラフ技術などと組み合わせるべき技術である、という見通しが共通認識として得られた。

■ 基調講演 「セマンティック Web (たとえば知識グラフ) はデジタル人文学で役立つのか？」

午後冒頭では国立情報学研究所の武田英明氏による基調講演が行われた。日本のセマンティック Web・知識グラフ研究を長年牽引してきた立場から、LLM全盛のいまこそ明示的な知識構造の意義はむしろ高まっている、という論点が提示された。

セマンティック Webとは、ウェブ上の情報に意味的な注釈を与え、人間だけでなく機械も解釈できるデータのネットワークとして整備していこうという構想である。その中核に位置するのが知識グラフであり、「主語・述語・目的語」という三つ組(トリプル)を単位として、「足利義満は将軍である」「将軍は天皇に任命される」といった関係を明示的なデータとして書き下していく。人物・場所・事物・概念にはそれぞれ一意の識別子(URI)が与えられ、別々のデータベースが同一の実体を指す際にURIを用いることで機械的に突き合わせることができる。結果として分野や機関を越えた情報連結が可能となる。

講演はまず、意味の表現形式を「単語集→辞書→シソーラス→タクソノミー→スキーマ→オントロジー」と段階的に整理するところから始まった。そのうえで、セマンティック Webの肝は技術そのものではなく、コミュニティ内部で暗黙のうちに共有されていた知識を外在化・明示化していく営為にあることが強調された。具体例として、明治期の百科事典『古事類苑』における引用関係のモデル化と、地震学における「岩盤のずれ」と「観測された揺れ」とを概念的に分離したオントロジー開発が紹介された。とりわ

け後者は、多数の地点からなる観測網が蓄積したデータを突き合わせるために、似て非なる概念を適切に切り分けることがいかに決定的であるかを示す好例であった。極めて強力だが本質的にはブラックボックスである LLM を、研究コミュニティが使いこなすための「結び目」として、以前から地道に続けられてきた知識構造化を位置づけるというメッセージは、後続する2つのディスカッションへの橋渡しを明瞭に果たすものとなった。

■ディスカッション2「一次資料と関連情報・コンテキストの接続」

続いてのディスカッションでは、江草由佳氏（国立教育政策研究所）、田中一孝氏（桜美林大学）、中川奈津子氏（九州大学）、Victor Eiti Yamamoto 氏（総合研究大学院大学）の4名が話題提供を行った。教育情報、西洋古典、方言の記録、移民や現代文化の資料群と、対象領域も資料の性質も大きく異なるケースが並び、そのことがかえって「接続」という問いの輪郭を際立たせた。

論点の第一は、一次資料のデジタル化が画像公開やテキスト化の段階で止まっているケースが依然として多く、関連情報と接続するために必要な構造化や典拠の付与に進めていないという実情の共有である。第二は、TEI（Text Encoding Initiative：人文学テキストの構造化記述のための国際的なマークアップ指針）、IIIF（International Image Interoperability Framework：画像資料をウェブ上で相互運用するための枠組み）、RDF（Resource Description Framework：知識グラフの基盤技術）といった国際標準を採用しても、コンテキストの粒度や境界は分野ごとに大きく異なり、共通のスキーマを上から押しつけるだけでは解決しないという指摘である。資料の特性を尊重しつつ相互運用性を担保する、精緻なモデリングが求められる。第三は、個別のプロジェクトが構築したデータを、相互に参照可能にするための具体的な枠組みである VIAF（バーチャル国際典拠ファイル：<https://viaf.org/>）や Wikidata（<https://www.wikidata.org/>）に接続することや、各プロジェクト間での URI の共有など、持続可能な運用体制の確立がなお課題として残っているという点である。質疑では、基調講演で示された知識の外在化をそれぞれの領域でどう実装するかが話題となり、AI を有効に機能させる前提としての資料構造化の重要性が確認された。



■ディスカッション3「歴史人物情報の構造化と共有」

最終セッションは、神崎正英氏（ゼノン・リミテッド・パートナーズ）、小川潤氏（東京大学）、グラムリヒ=オカ ベティータ氏（上智大学）、山田太造氏（東京大学）による話題提供で始まった。ジャパンサーチの RDF 化と人物情報の正規化、古代地中海世界を対象とする知識グラフの構築、近世日本を中心とした人物データベースである JBDB、東京大学史料編纂所の多様かつ多層的な人物情報の管理など、すでに相応の蓄積をもつ実践が紹介され、歴史人物情報を構造化・共有する際の具体的な課題が共通の俎上に載せられた。歴史人物情報を知識グラフとして整えるということは、個々の人物に固有の URI を与え、別名、生没年、役職あるいは関連する場所や出来事などを属性や関係として明示的に記述していくことを意味する。

論点として繰り返し語られたのは、人物同定の本質的な困難さである。同一人物が時代や場面によって複数の名を持つ例や、事典に掲載されていない僧侶や門人などを資料の読みから拾い上げる必要もある。その際に、ID として名前を使うべきか数字を用いるべきか、来歴に関する記録を併存させる理想的なモデルの実現可能性、AI による正規化の支援と人手による判断を組み合わせるヒューマンインザループ（Human-in-the-Loop）など、無数ともいえる選択肢の利害得失が率直に議論された。とくに、別名や襲名に伴う時間的・文脈的な差異をどう扱うかは、歴史研究の要諦そのものであり、単なる技術的課題には還元できない。プロジェクト間のデータ連携を、研究者が実際に扱えるかたちでどう実装していくかが、次の宿題として残された。

各セッションを貫いていたのは、「AI は使えるか、使えないか」ではなく「AI を使えるものにするための資料・知識の基盤をどう整えるか」という問いであった。本研究課題にて進められている TEI・IIIF・Unicode・知識グラフといった基盤整備の延長線上に、本ワークショップの議論は確かに位置づけられる。登壇者・参加者のみなさまに改めて謝意を表したい。

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。

2026年2月1日(日曜) [参加]

●第140回情報処理学会人文科学とコンピュータ研究発表会

<https://www.jinmoncom.jp/index.php?CH140>

情報処理学会の分科会である人文科学とコンピュータ研究会(<https://www.jinmoncom.jp/>)の2025年度冬季の発表会は、沖縄県石垣市のアートホテル石垣島で開催された。当日は天候こそ雨混じりであり、本土より格段に温暖な気候の中、予定をはるかに上回る34件の発表と多数の参加者を集め、盛況のうちに終わった。

プログラムは全日2会場平行で延べ8セッションに分かれ、多様な領域からさまざまなアプローチによる発表が並んだ。主題は、仏典・漢文・訓点・ヒッタイト文書などのテキスト研究を主とする文献学や古典学から、古今の歴史書・新聞・記録文書などを扱った歴史学、日本語・琉球語・ロシア語などの言語学、ゲーム・小説・脚本・アニメといった多彩なジャンルの文学・芸術学、アーカイブや展示の制作に関する図書館・博物館学、ほかにも社会学、考古学、教育学、メディア学といった幅広い分野に及んだ。

手法についてはとりわけ、近年ますます発達を続ける大規模言語モデル(LLM)をはじめとする生成AIを手法に取り入れた発表が、約3分の1ほどを占めた。AI技術の利用方法についても、入力・抽出(文字・画像認識およびマルチモーダル文書)、自動変換・要約、分析・評価(関連性、トピック抽出、品質評価)、出力・応答と多岐にわたる活用の試みがなされており、本研究の手法にとって参考となることも多かった。特に多言語・古典言語のテキストに対する既存言語モデルの適用事例について注目していたが、現状では人間の介入の余地がまだ大きいように感じられた。もちろん、AI一辺倒というわけではなく、やはり人文資料のデジタル化に重要なモデル化、構造化、アルゴリズムの探究に関する発表も多くみられ、さらに知識commonsの形成を論じたものも一定数あった。全体として、まさに現在進行中の新鮮で意欲的な報告が交わされる場となった。

なお、本研究課題からは、助成を受けた研究分担者・永崎研宣による蔵書印データベースおよびテキストデータ共有に関する2件の発表のほかに、複数のプロジェクト関係者が参加・発表を行った。

また、最後に国立歴史民俗博物館による企画セッションとして、歴史的建造物の3Dデータをデジタルアーカイブとして保存する取り組みの報告がなされた。翌2月2日にも、国立歴史民俗博物館などの主催により「時空間情報基盤の現在地」と題

する研究会が同会場で開催された。歴史的な暦・地図・行政区画などを表す情報基盤の構築成果について関野樹氏と北本朝展氏が解説する内容であり、多くの発表会参加者が出席した。

2026年3月13日(金) [共催]

Digital Humanities and Digitization Workshop for Japanese Materials

<https://www.nijl.ac.jp/en/news/10323/>

カナダのバンクーバーにおいて、3月中旬にアジア研究協会(Association for Asian Studies)が開催された。会議全体としては、北米のアジア研究者の集会として例年を若干上回る3,901人の対面参加者があり、大きな盛り上がりを見せた。北米のアジア研究者が一堂に会するこのイベントに並行して、毎年北米の東アジア研究司書による年次総会が開催されており、その一環として、プリティッシュコロンビア大学のロブソン・スクウェアキャンパスにおいて、「Digital Humanities and Digitization Workshop for Japanese Materials」が開催された。これは本事業をはじめとするいくつかの事業による共同開催のイベントであり、その内容は、本事業でも中心的なテーマとなっているTEIガイドラインを中心としたテキストデータ構造化に関するセミナーであった。主な講師は研究分担者の永崎がつとめ、英語による3時間にわたる講義と実習が行われた。北米の日本研究司書や日本研究者を中心に45名ほどが参加し、講義と実習の後には闊達な議論が行われた。

2026年3月21日(土) [共催]

人文学のためのテキスト構造化ワークショップ

<https://kemco.keio.ac.jp/all-post/20260321/>

慶應義塾大学三田キャンパスにおいて開催されたこのワークショップでは、文部科学省委託事業「人文学・社会科学のDX化に向けた研究開発推進事業(JPMXP1624)」(再委託先:慶應義塾/実施機関:慶應義塾ミュージアム・commons(KeMCo))が作成に取り組んできた、東アジア・日本語テキストの構造化に関するガイドラインと『手引き』についての検討と、東京大学史料編纂所における日本史史料集の構造化フローの整備についての報告が行われた。このガイドライン及び『手引き』は、現在はWebで(<https://teikem.kemco.keio.ac.jp/eajguidelines/>)公開されているものであり、ガイドラインは、東アジアのテキストデータ構造化に関する基本的な考え方についての解説となっており、『手引き』は、テキストデータの構造化の考え方に加えて、それを具体的にどのようにして実践するか、という観点から様々な手法を紹介している。オンラインも含めると88名の参加申込みがあったこのワークショップでは、これらの原稿の最終稿に近いものが共有され、その改善案について様々な角度から議論が行われた。この内容には本事業の成果も反映されており、これを踏まえた本事業のさらなる発展が期待される。

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