

パネル「慶應義塾大学 HUMI プロジェクトのデジタル・リサーチ・ライブラリー構想」

パネリスト 高宮利行 (慶應義塾大学)
岩井茂昭 (慶應義塾大学)
松田隆美 (慶應義塾大学)
アンドルー・アーマー (慶應義塾大学)

慶應義塾大学が構想するデジタル・ライブラリーは、図書館が所蔵する稀覯書コレクションをデジタル画像化して、人文科学諸分野の研究者、学生の研究利用に供することを目的としている。この構想を実現するべく、1996年4月に準備研究組織として Humanities Media Interface Project (通称 HUMI プロジェクト) を発足させた。この組織には文学部、理工学部、環境情報学部などから研究者が参加しており、相互に情報を交換し合いながら、稀覯書のデジタル化の作業を進める一方で、人文科学的コンテンツとデジタル技術の理想的な融合形態、さらにネットワーク利用を含む検索、公開の手段を模索している。

Panel: Digital Research Library Envisioned by the HUMI Project at Keio University

Panelists: Toshiyuki Takamiya (Keio University)
Shigeaki Iwai (Keio University)
Takami Matsuda (Keio University)
Andrew Armour (Keio University)

One of the primary functions of the digital library envisioned by Keio University would be to provide researchers and students in the humanities with digital images of rare books held in Keio University Library. It was with this in mind that the Humanities Media Interface (HUMI) Project was established in 1996 with the participation of members of three different faculties — Literature, Science & Technology, and Environmental Information. Through joint experiments in the digitization of manuscripts and books, we are currently developing a prototype for a future digital library — offering networked access to scholars worldwide — in which advanced technology complements traditional textual studies.

1. デジタル書物学に向けて (高宮利行)

書物に書かれた内容については、文学はもとより人文科学、自然科学を問わず、あらゆる学術分野で研究対象とされてきた。ところが書物の内容とは別に、書物の外面的な成立過程について研究する学問、すなわち書誌学がイギリスで誕生したのは、わずか 100 年ほど前のことであった。1877 年に、ウィリアム・キャクストンがウエストミンスターに印刷所を構えて 400 年を経たのを記念する事業があり、そのころから活版印刷術やそれを用いた印刷テキストの諸問題を追究する書誌学が成立してきた。

もっともこの行事自体は、19 世紀半ばのドイツでのグーテンベルクによる印刷 400 年祭や宗教改革 400 年記念行事がナショナリズムによって触発されたように、ヴィクトリア朝の大英帝国の国威発揚の気運と結び付いていた。

ロンドンに書誌学会が生まれ、機関誌 (現在は『ザ・ライブラリー』) に精緻な研究成果が発表されるようになると、イギリス流の書物の物理的な外面相に関する研究方法は、ほどなくドイツやアメリカに飛び火した。しかし、中世には写本が多数生産され、インキュナビュラ (15 世紀の初期印刷本) が多く出版されたはずのフランスやイタリアでは、イギリス流書誌学はさほど定着しなかったといっている。

書誌学が重箱の隅をつつくような個別研究、たとえばエリザベス朝演劇の印刷本文に見られる I will や I shall の短縮形に対する I'll, Ille, I'le といった異綴から何人の植字工が同時に仕事をしてきたかを分析する、といった研究によって支えられていた 20 世紀半ばから、フランスではアナール派の歴史学者たちが、その方法論として書物史の観点を用いてめざましい研究成果を発表していった。一般大衆に迎えられた印刷本がいつ、どのような状況で出版され、いかなる読者層にいかにか受容されたかを社会学、社会史的に扱おうとする書物史の領域からは 18 世紀フランスの様相を中心に、優れた研究書が世にでた。アメリカに広まった書物史の方法論は従来日の目の当たらなかつた 19 世紀アメリカの大衆雑誌やダイヤモンドヴェルなどにも応用され、ペンシルヴァニア大学には書物史学部も誕生した。

コンピュータを用いた書物の研究も盛んになった。当初は統計的な処理をするのに用いられていたコンピュータは、その後文学研究者によってコンコードランス作成に利用されるようになった。

コンピュータの性能が上がり、大容量を必要とする画像処理が可能になると、従来は光学的な手段によって研究されてきた損傷の激しい写本の画像分析に用いられるに至った。そのよい例が大英図書館が所蔵する『ベーオウルフ』写本である。イギリス最古の叙事詩の現存するただひとつの写本は 17 世紀に火事にあつたこともあつて、判読が困難な箇所が多い。折しも 1980 年代初頭から作品と写本の制作年代をめぐる議論が百出した。これを解決する手だてのひとつとして、写本のすべてのページをデジタル画像化する計画が始まり、今日ではインターネットを通じたアクセスが可能になっている。これが人文学における本格的なデジタル書物学の誕生といつてよい。現在では『死海文書』なども、デジタル画像化されて CD-ROM で利用できる。

このように書物に見られるテキストとイメージの関係も最近とみに注目を浴びるようになった。われわれ HUMI プロジェクトが目指すのも、書物の内容のデジタル画像化である。すでに慶應義塾図書館が所蔵する「グーテンベルク聖書」については、この 3 月に NTT の技術支援を得て全ページのデジタル化に成功した。これによって、図書館の資料を内外の研究者の利用の便に供し、現物を損傷することなく次の世代に継承させるという、従来貴重書図書館が抱えていた難問を解決できる糸口を得た。今後の詳しい計画と諸問題については、各発表者に譲りたい。

2. 稀観書のデジタル入力に関する諸問題 (岩井茂昭)

HUMI プロジェクトにおけるデジタル化は、慶應義塾図書館が所蔵する稀観書コレクションを対象として行われている。西洋中世の写本、ルネサンス期の初期刊行本、近代の図版入り博物誌資料、和漢書、浮世絵などのデジタル化作業が進行している。

これらの稀観書は装丁およびそこに記録されている文字と図版の形態に資料価値があるため、テキスト・ベースではなく画像ベースでデジタル化する必要がある。研究者がディスプレイ上にその画像を表示したときに、すべての文字の形態や装飾、図版などを詳細に眺めることができないなければならない。

このようなデジタル化方法を実践する場合、書物の物理的な形態と入力条件が顕著な障害となる場面がいくつか見受けられる。主要な問題を列挙すると以下のようになる。

- (1) 破損の危険性
- (2) ページ面の湾曲
- (3) 入力時間
- (4) 色彩の再現性

(1) の「破損の危険性」は数百年を経て現代に伝えられる書物を取り扱う上で常につきまとう問題である。現代の書物をテキスト・ベースでデジタル化するのであれば、フラットベッド型のスキャナに押し付けて入力したり、あるいは装丁をばらして入力することも考えられるが、文化遺産としての価値を持つ書物をこのように取り扱うことはできない。古い書物の場合には一定以上の角度でページを開くと装丁が破損してまう心配があり、また、これは (2) の「ページ面の湾曲」という問題とも関わりがあるが、ガラスなどをページ面に押し付けたり、真空吸引などを行ってページの平面性を出そうとすると、インクや顔料などが剥落してしまう可能性が高い。このような危険を回避するためには、カメラ型の入力装置を用いて書物に最も負担をかけないポジションを確保しながら慎重な入力作業を行わなければならない。

(2) の「ページ面の湾曲」は、装丁のもろい書物のページ面を一定以上の角度で開くことができない場合あるいは対象書物のページ数が多い場合に問題となる。ページ数が多い場合には装丁の破損を心配する必要がなくとも、開いたページの装丁側のノドの部分が湾曲し、空白部分が少ない場合には文字や装飾の形に歪みが生じる。これをこのままの形で入力すると、ディスプレイに画像を表示したときに何が映っているのか判別できなくなってしまう。この問題を回避するためには入力時にページの平面性を最大限に高めるための工夫をすることに加え、三次元画像処理によって湾曲を平面にする手段が考えられる。

(3) の「入力時間」は一定数量以上の書物をデジタル化してアーカイビングしようとするときに、どのような入力器材を用いるべきかということに直接影響を与える。上記 (1) の制限があるためカメラ型の入力器材を用いることになるが、目的に応じて最も適切な入力器材を選択しないと、大部の書物のデジタル化に必要以上の時間と経費をかけることになったり、逆に、効率のみを考えて必要な精度の画像を確保できないという問題が生じたりする。今後デジタル・リサーチ・ライブラリーを構築していく上で、入力器材として何を選択するかは最も大きな問題のひとつといえるだろう。

(4) 「色彩の再現性」の問題も入力器材の選択と直接関わりを持っている。デジタル・リサーチ・ライブラリーを想定したデジタル化では、最終出力がディスプレイとなる可能性が高いが、入力器材自体の色彩再現性能や入力時のライティング条件、閲覧用ディスプレイの性能などを考慮しなければ、リサーチ目的にかなうデジタル画像を提供することはできないだろう。

3. 「知の遺産としてのデータベース」(松田隆美)

文学、歴史、美術史など、世界各地に分散する多種多様な一次資料を研究対象とする分野にとって、文字資料のフルテキスト・データベースの作成、電子図書館計画といった近年の動向が、情報のグローバルなアクセスと検索を可能にすると言う意味で、重要なことはいまでもない。しかし、たとえば一千年の歴史を持つイギリス文学の主要作品すべてを、一枚の CD-ROM 上で検索し読むことが可能であるということは、これまでにない通時的なひろがりを持った新たなリサーチの可能性を提供すると同時に、個々のテキストを、それが本来持っていた歴史的な脈から切り離し再構成することに他ならない。テキストが歴史の産物である以上、それが後世へと存続してゆくこと自体が元来の歴史的な脈からの切り離しに他ならないが、すべてのテキストを単一フォーマットの電子テキストに変換する大型のフルテキスト・データベースの存在は、同じことをより徹底したかたちで容易に実現している。人文科学においてこの種の全文データベースの製作はすっかり一般的になったが、その一方で電子化の短所を補い、コンピュータの寄与を新たな局面において展開してゆくことが必要である。

具体的には、人文科学系の研究者が個人あるいは数名の単位で遂行可能なプロジェクトとして、文字資料のデジタル化とインターネットを利用したオンライン・パブリッシングが挙げられよう。デジタル化は、書物を表紙から裏表紙まで全ページを画像としてデジタル化して、電子テキストに還元することなく、その書物が製作され実際に利用されていた時代の形態を損なうことなく伝えようとするものである。対象となる書物としては、たとえば中世、ルネサンス期において広く使用された大型の事典類が考えられる。これらのレファレンスブックは、書物自体が大部であるため全文の電子化は現実的ではなく(近代初期の活字は、組文字や略号などを多用しているため OCR による処理も容易ではない)、また印刷によるファクシミリや学術的校訂版も作成されていない。その結果、書物自体の重要度にもかかわらず、欧米の一部の図書館のみが所蔵する稀覯書となっている場合が多い。デジタル化によるファクシミリの製作は、そのような書物を可動性の高いかたちで提供することにより、たとえば 16 世紀ヨーロッパの著作家のレファレンス・ライブラリーを再現し、個々のテキストを本来の知的な脈のなかに位置づけることを可能にする。

さらに注目すべき点は、当時広く利用された百科事典や命題集の類は、主題別の索引を何種類か備えており、また本文自体にも詳細なクロスレファレンスと欄外脚注を備えている点である。この検索システムをソフトウェア的に再現することで、たとえばルネサンス期の知識の体系化と認知の方法を追体験できよう。デジタル化を通じて研究者は中世、ルネサンスの知的生産作業を往時のままにヴァーチャルリアリティとして体験し、さらにデジタル化の特性をいかして情報を新たな視点から検索、分類することで、知的認知のありかたを膨らませてゆける。

中世、ルネサンス期の書物の詳細なクロスレファレンスのシステムは、テキストが常にインターテクスチュアルな存在であることを再認識されてくれるが、古い書物であればあるほど、その意義と評価は時代や地域によって変化して今日に至っている。特定の書物の相対的な位置づけをリアルタイムで反映できる書誌学的道具として、インターネット上のオンライン蔵書目録が考えられる。特定の書物に関する引用にリンクを張り、情報とリンクを常に更新することで、最新の学術的成果を常に公開してそれによって自らの位置を修正できるだけでなく、テキストの意味を常に更新し続けている我々の時代の知的な脈を体験することができる。

近年の歴史学や文学研究の過去を分析する視点は、特定の事件やテキストを見つめ、それを理解する時代の視座を重視する。デジタル化とインターネットの応用を通じて得られる道具は、個々のテキストに相対的意味を与える時代の知的な脈をとらえる道を開いていると考えられる。

4. 蜘蛛巣城：電子図書館の実現 — ホメロス卷子写本（偽作）の例（アンドルー・アーマー）

「アレクサンドリア図書館から奇跡的に救われた」と伝えられるホメロス卷子写本の偽作を具体例として、慶應義塾 HUMI プロジェクトにおけるデジタル・データの流れを追いながら、その過程において見出される諸問題を取り上げ、将来の電子図書館のあり方を探る。

今回の研究対象であるホメロス卷子写本は、およそ 50mm という極めて狭い幅に、細かい文字がほぼ 1mm 間隔でびっしり書かれており、また羊皮紙の状態も良いとはいええないため、ルーペなしではほとんど読むことができない。このような特殊な資料の場合、デジタル画像処理による判読性の向上が必須である。今回は大型アナログカメラで撮影したポジフィルムを Kodak 社の PhotoCD スキャナを用いてデジタル化し、その画像を Adobe 社の PhotoShop によって処理してみた。この写本の場合には薄茶色の羊皮紙の上に成分不明の茶色いインクで細かい文字が書かれているために、拡大と文字の色の変換が必要である。この例のように、どのような方法でデジタル化しどのような画像処理を施すことによって最も効果的に判読性を高めることができるのかという知識の蓄積は、電子図書館が提供すべきサービスの形態を考える上で重要であろう。

ホメロス卷子写本が提示する別の問題は文字の並び方と使われている文字の形態である。この写本では牛が畑を耕す時の歩き方にちなんだ *boustrophedon* (犁耕体) という古代の書式が使われており、一行一行を交互に逆の方向に書いて行く。そして、使われているギリシャ文字も通常の形とは異なっている。研究利用するためには、これらをすべて単一フォーマットの電子テキストに変換する必要がある。その場合、特殊な書式に関する知識の蓄積と、どのような文字セットを利用すべきかという方針の確立など、電子図書館が解決しなければならない新たな課題が浮かび上がってくる。

今まで述べた個々の問題を解決し、やがて画像処理されたイメージ・データと電子テキストをどのような形で研究者に提供すべきであるのかは、これからの電子図書館のあり方を考える上で重要な問題である。慶應義塾大学 HUMI プロジェクトでは、イメージ・データとテキスト・データをインターネット経由で提供するという基本姿勢のもとに、Web サイトの構築を目指しているが、上記のような問題は、インターネット上での情報発信の方法も大きく左右する。

古代ギリシャの文字と犁耕体、巻き物の形状等を扱うためには既存の HTML だけでは間に合わない。もちろん、刻々と変化し続ける WWW の世界では、数多い plugin のほかに、Java や ActiveX、Microsoft 社と Netscape 社がそれぞれ推進している二種類の Dynamic HTML をはじめとする新しい試みが進行中である。このような環境下、情報発信の方法として電子図書館がどの道を選択していけばよいのかが重要な課題となる。たとえば、巻き物を巻き物らしく見せるには、上下左右自由にスクロールあるいはズームできる機能が必要ではあるが、あまり普及していない plugin に依存すれば、アクセスできるユーザーの数が制限されて、本来の電子図書館の使命を損なう結果になりかねない。現在のところ、Java を利用する表示システムが有効であると考えられる。

インターネット上での情報発信の手段の選択は、当該資料を用いた研究の発展に直接的な影響を与える可能性がある。例えば、HUMI プロジェクトのメンバーの中には、このホメロス卷子写本の古代文字を正確に解読できる研究者はいないため、国内外の専門家に意見を求める必要があるが、当方の用いるネットワーク上の資料表示方法システムに一般性がなければ、期待される反応を得ることはできないだろう。この問題は、電子図書館を中心とした学術的なネットワークを構築する上で重要な課題を投げかけている。現時点では WWW 上での情報発信に関し、さまざまな手段が乱立しているが、特定の研究目的を共有するコミュニティの間で情報発信・受信の手段のプロトタイプが確立しなければ、該当分野における研究の発展は困難になると思われる。