

哲学・思想研究における人文情報学の可能性

永 崎 研 宣

はじめに

紙媒体による情報共有を前提とした学術研究活動としての人文学の歴史は長い。哲学・思想研究においてもそのことは例外ではなく、むしろ紙媒体を通じて交換されるテキストを通じて学問を発展させてきたという点において、まさに紙媒体に大きく依拠して展開されてきたと言ってよいだろう。しかしながら一方、デジタル媒体が当たり前用いられ、むしろ紙よりもコストの低い情報共有手段として流布するに至った現在、紙媒体で行われてきた研究活動もまた、デジタル媒体へとその軸をある程度置かざるを得ないことになる。下田正弘『仏教とエクリチュール』¹が述べるように、このことはオングが指摘した音声メディアから書写メディアへの移行によってはじめて音声メディアの構造的な反省が可能となったという枠組み²に類似する面がある。すなわち、デジタルメディアの登場（発明としての登場ではなくコモディティ化したという意味で）によって紙媒体で行われてきたことが反省的に捉え得るものとなったということであり、それを通じて、これまでの学術研究活動が紙媒体を通じてどのようなことを行ってきたかということが明らかになるとも言えるのである。

もちろん、ここでの紙とデジタルの差異は慎重に捉えられなければならない。なぜなら、デジタルメディアのいくつかの側面は、紙媒体の特徴と異なるものではなく、むしろ紙媒体の変奏曲としてとらえるべき面もあり、いわば、アイゼンシュタインが言うところの印刷革命³の延長試合、もしくはその本戦を継続しているに過ぎないということかもしれないからである。これは、その外側の立場を取ることが難しい現在の我々には十分に検討できない課題である可能性があるが、それでも、この点には十分に配慮していく必要がある。

このようなパースペクティブを踏まえつつ、本稿では、人文情報学（Digital Humanities, デジタル人文学）が哲学・思想研究にもたらし得る可能性のうちのいくつかについて検討してみたい。

人文情報学との様々な接合面

哲学・思想研究と人文情報学が接合し得るのは一つの側面だけではない。まず思い浮かぶものとして、近年改めて脚光を浴びつつある AI（人工知能）との関わりが挙げられる。一頃は哲学研究においても AI が大きなテーマとなっており、その頃にいったん大体の議論は尽くされたように思われる⁴。技術的な限界の到来とともにいったん収束した AI ブームは、近年、深層学習と呼ばれる手法として新たに注目されるようになり、具体的な産業に結びつく場面も出てくるようになってきている。一方で、理念的な議論としては 2005 年に Ray Kurzweil が提唱したシンギュラリティ（AI の能力が人間の思考能力を超える技術的特異点）⁵ を受けて議論がより具体化してきており、宗教を扱う分野においても AI をテーマとした議論が散見されるようになってきている⁶。しかしながら、哲学・思想分野で主要なテーマとしてよく扱われてきたこの課題には本稿では踏み込まず、むしろ、この分野も含めたもう少し広い人文学、とくに、文献資料を扱う人文学全般にとっての接合面を扱っていききたい。

このような接合面については、人文学の研究活動という観点から検討してみると、大別して 2 つに分類できるように思われる。それは、「研究基盤の構築」と「気づきの支援」である。これらについて以下に少し検討してみよう。

「研究基盤の構築」

「研究基盤の構築」とは、デジタル技術の効果的な応用を前提とした研究基盤、すなわちデジタル学術研究基盤の構築に関わる事柄である⁷。デジタル技術が導入される直前まで、人文学の研究基盤は紙媒体を軸としたエコシステムによって成立していた。研究者が調査した文献資料やそれに基づく先行研究としての書籍は図書館や文書館等に集積されており、そこに行けば読むことができる。研究対象となる資料はそれぞれの文化的コンテクストに基づいた構造によって記述されているが、それに基づく 2 次資料は章・節といった階層構造によって記述され、脚注や章末中など、いくつかの形態に従って内容が記述され、我々はその構造を、多くの場合は特にガイドもないままに理解しながら読むことができる。デジタルがないころであれば、その資料を探すのは図書カードからということになるが、それも図書カードの基本的な記述項目がわかればそれで探索可能である。それらを調べつつ必要に応じてメモやノートを作成し、それらを踏まえて論文や書籍を執筆する。執筆した原稿は学会の編集委員会や出版社に手渡され、しかるべき手続きの後

に刊行される。刊行されたものは、学会であれば会員に郵送されたり図書館に所蔵されたりする。さらにそれは国立国会図書館にも納本される。そして、それは次の研究をする人によってまた参照され、次の成果を生み出していくことになる。このプロセスの中で、人文学の研究者の多くはあまり関わることのないプロセスがいくつか存在し、人文学研究者としての現在の誰かが設計したわけでもないのに自然とできているプロセスもある。しかしそれらはいずれも、自然とできたわけではなく、誰かが作り、流布や模倣などによって誰かが広めたのである。日本の人文学のほとんどは明治期に西洋先進諸国からある部分を輸入したのであり、その時点ですでにある程度完成されていた部分もあっただろう。個々のどの部分が西洋発でどの部分が日本由来か、ということについては、中国文化圏からの移入も考慮しなければならず、本稿の手に余るため、割愛させていただく。ここで重要なのは、ほとんどのことはどこかで誰かが決めていたということである。探しに行くとき用意されている図書カード、そして図書カードに書くべき事柄。書架で本を並べる順番。本の冒頭には序文があり、目次があって本文がある。本文のページのどこかの隅に書かれた数字は頁番号であり目次の各項目に付された番号と対応している。本文中の文字の右上に付された数字は脚注の番号であり、これに対応する番号の脚注はその箇所へのなんらかの説明や異文などを示している。その内容がなんであるかということについては、脚注の内容を見て読者が判断することになる。また、本の最後には索引がついていることが多く、その各項目にもやはり頁番号が付されている。そのようにして読んだ本や資料に基づいて書いた原稿は、原稿を刊行してくれる担当者に手渡され、校正等のやりとりがあって最終的には多くの場合刊行される。他にも様々なプロセスがあり、ここで挙げたプロセスもまた細部に注視すれば様々な部分が抽出できるだろう。今回、まず着目しておきたいのは、個々のプロセスではなく、このような個々のプロセスの多くを哲学・思想研究者を含む人文学者は自らそのスタンダードを作ったわけではなく、ただ受容し、場合によってはそこから派生する不便さに甘んじてきたという点である。デジタル技術が導入されたことにより、こういった一連のプロセスの多くは作り直しを余儀なくされた。その作り直しの過程においては、デジタル技術の可塑性や変化の速さのため、一定の介入の余地が不定期に生じている。すなわち、これまでは所与のものとして研究活動の側が時として無理にあわせていた研究基盤が、研究活動のニーズに応じて改良することが可能となってきたのである。とはいえ、個々の研究者が一人で対応することは容易ではなく、研究基盤の構築に関連する既存の団体を通じて意向を伝えたり、似たような志向を持つ研究者等によるコミュニティを形成してニーズについてすりあわせを

行った上で対処したりすることが、比較的实现性の高い方法になってきている。哲学・思想研究においても、後述する「哲学者のためのテキスト・データベース研究会」や宗教研究における聖書や仏典等のデータベースの構築、そして、紙媒体と異なり簡単にコピーと再配布ができてしまうそれらのものを適切に共有するための法律や制度の整備、もはや商業出版物のような集金システムが適用できずこれまでの評価軸もうまく適用できないそれらの成果物を評価するための手法、実際のデータベースの使い勝手やそこに組み込むべき様々な成果とその組み込み方、そしてそこから生み出される様々な統計結果の扱い方など、紙媒体の時代にはなかったことや存在したとしてもあまり注目されなかったような様々な事柄が、新たに設計し直すことが可能なものとして現前しているのである。哲学・思想研究のコミュニティ単独では難しいような事柄であっても、人文学における他の分野、歴史学や文学、人類学、心理学、社会学など、様々な分野と協働することで対処可能になることもある。後述するTEI (Text Encoding Initiative)⁸はまさにそのようなコミュニティとして30年間以上その活動を継続している。デジタル以前に長い時間をかけて設計され普及してきた様々なプロセスが、社会全体の情報流通の仕組みの大規模な変革という状況の下で再設計を迫られるなか、哲学・思想研究が自らのあり方をよりよく社会に示していくためにできる選択肢には様々なものがある。いずれも簡単にできるのではなく、取り組むべきことすべてに取り組むだけの資源はもはや持ち合わせていないと思われるがゆえに、何が重要になり得るかをよく見極めた上での集中的な資源投下が必要だろう。

「気づきの支援」

ここで意図する「気づきの支援」とは、研究活動において何らかの新たな発見を得る契機を「気づき」という言葉で表現した上で、情報技術を活用することによってそれを得るための支援を得ることを指している。ここでは、単なる全文検索からテキストマイニング技術やディープラーニング技術を用いた様々な分析に至るまで、様々な手法を念頭に置いている。あるいはまた、哲学・思想研究にかかる教育における有用性を志向する取り組みもある⁹。このような情報技術による分析は、基本的には客観的な結果を入手できるものの、元になるデータや分析手法が十分に適切なものでない場合もあるため、結果に対して全面的に依拠することが難しい場合もある。いずれにしても、その結果にどういう意味があるかを判断するのは、そこに何らかの研究上の文脈を見出すことのできる研究者が行うということになる。研究上の文脈や分析対象となるデータ、分析手法、分析プログラム等に応じて分析結果への

依拠の度合いは様々だが、人が目を通すことのできない規模のデータが研究対象とならざるを得ない分野においては多かれ少なかれこの手法に依拠せざるを得ないことになるだろう。人文情報学分野では、このような状況に対応するために遠読 (Distant reading)¹⁰ という手法が提供されて久しい。哲学・思想研究においても、何らかの形である程度はそういった手法を採り入れていかざるを得ないだろう。

「気づきの支援」は、「研究基盤の構築」とある面で非常に密接な関係がある。というのは、気づきの支援をより良くするためには、気づきを得るための資料や手法を供給する研究基盤そのものを改良することが有効だからである。人文学研究全般にせよ哲学・思想研究にせよ、それぞれのニーズにあわせた分析対象や手法があり得る。個々の研究者のレベルでも着目点は実に多様である。上述した研究基盤構築への介入とは、単にプロセスに参加すべきであるという規範意識に拠るのみでなく、技術に応じて研究を発展させるための手段と捉えることもできるのである。情報技術は発展がはやく、それを研究に応用しようとしたとしても、具体的な技術の多くは徐々に廃れていってしまう。しかしながら、あまり廃れない部分も一方では存在し、そこに注力するならば労力の無駄を心配する必要もかなり減らすことができるだろう。その点については後述するが、以上のような接合面を踏まえつつ、次に、人文情報学のこれまでの経緯について、以下に概観してみよう。

人文情報学の経緯

黎明期の動向

本稿で人文情報学と呼んでいる学問領域は、1940年代半ばにその淵源をたどることができるとされている。ロベルト・ブサ神父が1946年に構想したトマス・アクィナスの電子索引 Index Thomisticus がそれにあたる¹¹。ここで出てくるのが哲学・思想研究に関わるものであるということは、コンピュータが知識や知能に関わるものになっていくことをまさに象徴する出来事だったと言えるだろう。1949年にはIBMが支援を表明して作成が開始されるが、当時のコンピュータの性能や情報処理技術の特殊性など、乗り越えるべきことも多く、最初に刊行されたのは1974年のことだった。この間にも、国内外で徐々に情報技術を人文学に応用するための様々な取り組みが国内外で進められていく。1950年代末には計量国語学会においてコンピュータを利用した研究が発表されており、さらに、1973年にはロンドンにおいて現在の欧州デジタル・ヒューマニティーズ学会（当時の名称は文学・言語

学情報処理学会、Association for Literary and Linguistic Computing)¹²が、1978年には米国にて Association for Computers and the Humanities が設立される¹³など、学会が設立されるほどにこの種の研究が広がりを見せつつあったことがみてとれる。テキストの量的な分析から新しいことを発見しようとしたり、既存の研究成果を数量的に裏付けたりしようとする研究が各地で展開されていた。また、分析とまではいかずとも、資料の目録を利便性の高い手法で作成共有するために情報技術が用いられるようになりつつあった。

このような流れを受けて生じた一つの画期が、TEI (Text Encoding Initiative)¹⁴である。1980年代初頭にIBMがパーソナルコンピュータを大きく普及させることになるPC/AT互換機を発売するなど、コンピュータの利用が人文学研究者にも手の届くところに来たことから、電子的な人文学資料であるところの電子テキストも広く行われるようになった。しかしながら、各地で別々に作成された電子テキストは、互換性が不十分であり、また、効率的な作成の仕方も共有できていないという状況が徐々に明らかになっていった。その問題を解決するために、複数の研究助成団体の支援を受けて1987年に開始されたのがTEIであった。

ニューヨーク・ポキプシーに集まった人文学者と情報学者達には、米国のみならず欧州や日本からの参加者も含まれていた。議論の末に、ポキプシー原則としてまとめられ、これに沿った人文学テキスト資料のための電子化ガイドライン、すなわち、TEIガイドラインが作成されることとなった。

人文学資料のために電子化し、それを何らかの有用な構造として作成しようとするということは、人文学がその活動において資料中に見出す内的な構造について十分に理解している必要がある。そして、人文学と一口に言っても様々な分野があり、テキスト資料を扱うものに限ったとしても、哲学、歴史学、文学をはじめ、それぞれが資料に対する多様なアプローチを含んでいる。それゆえに、電子化ガイドラインは、人文学における個々のアプローチを熟知する当事者たる人文学者達の手によって策定が進められることとなった。このことは、人文学における電子化の方針を策定することが、特定ドメインを対象とした情報構築についての研究という、一つの研究領域を形成していくことをも意味した。それまでは徹底して紙媒体の制約を活かすことで展開されてきた研究資料が、その制約を離れた時にどのような形であり得るのか、そしてそれによって研究がどのように変化していくのか、ということは、思弁的な検討課題であり得ると同時に、TEIという営みを通じてそれを実践的に検証することも可能となったのである。

しかしながら、TEIの登場した時代は、未だコンピュータの利用が広く普

及していたわけではなくコンピュータ利用自体の技術的制約も強く、これがその意義を大きく花開くには、今少しの時間を必要とした。すなわち、コンピュータの性能向上とインターネット・Webの普及という技術面での進歩と、Unicodeの発展と普及による多言語環境の容易な実現という大きくは二つの要素によって、日本も含め国際的に情報の交換が極めて容易になることが、TEIに潜在する本来の可能性を発揮するためには必須であった。この環境が実現していくなかで、人文情報学自体も有機的にその姿を変化させていったのである。

日本における前駆的状况

この時期、哲学・思想研究との接点という意味で注目しておきたいのは、1980～90年代に日本で進められたテキスト・データベースの構築である。1985年9月3日、東京大学大型計算機センター集会室において第一回の研究会が開催された「哲学者のためのテキスト・データベース研究会」¹⁵のプログラムを見ると、哲学研究のためのテキスト・データベースを構築するための基礎設計や実際についての発表が5件行われていたようである。第3回目からは「哲学者のための」という名称が外れ、広く人文・社会科学研究者のための「テキスト・データベース研究会」となり、英語名は「JACH (Japanese Association for Computer and Humanities)」とした旨、『JACH ニュースレター』第一号(1986年9月18日刊)にて報告されている。おそらくは、米国のACHを受けた名称と思われる。この段階では、コンテンツとしては、デカルトの『省察』、ギリシャ語原典、トーマス・マン、東洋学が挙げられており、技術面では、テキスト・データベースの各種機能やOCR、人工知能があげられる一方、その楽しさや課題についての発表も行われていた。その後、研究会で扱われる話題は徐々に幅を広げていき、キーワードを拾い上げてみるなら、プラトン索引、科学論研究、解釈学、漢字字体、中世英語英文学、ウィトゲンシュタイン、国語調査、漢文テキスト、和漢文OCR、LaTeX、美術史、デーヴァナーガリー文字、スタインベック、アラビア語テキスト、G.H.ミード、英日自動翻訳、国文学研究、源氏物語、タイ語テキスト、中世朝鮮語、チベット語、太宰治、西ドイツ中世史、といったものがみられた。

1991年6月28日の第13回研究会では、当時TEIを推進していた研究者の一人であるオックスフォード大学のSusan Hocky氏による講演とともにTEIが主に取り上げられサンسكريット・文学研究・日本文学研究・マルチメディアへの適用可能性が議論されたようである。筆者が入手した資料ではこの次の1991年12月13日の仙台国際センターでの開催を最後に記録が途

絶えている。この会では、インド学仏教学からコンピュータに取り組む6件の発表とTEIに関する発表が1件という構成であった。

JACHが哲学系テキスト・データベースを志向して始まっていたのに対して、1987年頃より日本文学テキストのデータ共有を目指した情報処理語学文学研究会も運営されていた。こちらは英語名をJALLC (Japanese Association for Literary and Linguistic Computing)としており、JACHが米国の人文情報学系学会の名称を受けているのに対して、こちらは欧州のものを受けていたとみられる。哲学・思想研究的な発表が扱われたのは仏教学に関するものが散見されるくらいであったものの、テキストデータを共有するという点に関しては深い検討が行われていた。1989年に始まるGNU GPLによるコピーレフト運動の影響も強く受けたようであり、現在のオープンデータ推進の先駆的な活動であったと言ってもいいようなものであった¹⁶。ここでもTEIに関する発表が散見され、検討が行われたことは見てとれるが、一方で、文字コードの問題やSGML時代のTEIの課題なども指摘されており^{17,18}、当時の日本で導入することが容易ではなかったことが見て取れる。

なお、JALLCの活動は、研究者同士でのテキストデータの共有を目指す相互的なものを基本としていたようであるが、筆者が確認できた限りでは1998年7月18日を最後としている。

このように、JACHやJALLCの活動により、デジタル時代の研究基盤としてのテキストデータを構築しようとする営みが日本においても着実に行われており、さらにそこにはTEIガイドラインを導入しようとする動きもあったようである。このことは、情報知識学会(1988年設立)のニューズレターにおいてもその片鱗を垣間見ることができる。その設立記念号ではJACHの活動紹介が掲載されており¹⁹、当時の活発さを想起させるものがある。その後、哲学研究者としてテキスト・データベースに熱心に取り組んでいた長瀬真理氏によりTEIに関する報告も行われていた²⁰。

また、研究集会としては、京都大学大型計算機センターの研究セミナーとして1990年3月に初回が開催された「東洋学へのコンピュータ利用」研究セミナーにも言及しておきたい。これはテキストデータベースというよりはデータベース全般ということで東洋学におけるコンピュータ利用に関する様々な取り組みが毎年報告されてきており、これは現在でも毎年一度、活発な研究会が開催されている。

あるいは、こうした学会・コミュニティ活動からはあまり見えてこないものとしても、哲学・思想研究においては、様々な哲学者のテキストが入力されて活用されており、フッサールやヘーゲル²¹のテキスト・データベースが日本で作成され、関係者の間で活用されたようだが、著作権問題で公開が十

分にできないという例²²もあったようである。宗教関連文献でも枚挙に暇がないが、国内の例を挙げてみるなら、たとえばサンスクリット仏典の全文テキストデータベースに基づいて作成された「ダルマキールティ梵文テキスト KWIC 索引」²³では刊行後に元になったテキスト・データベースも Web 公開された。あるいはまた、パリー語聖典のデータベース化への取り組み²⁴、1994 年の SAT 大蔵経テキストデータベース研究会による大正新脩大蔵経への着手など²⁵、この頃には大規模なものへの組織的な取り組みも行われるようになっていた。このような流れの終わりの時期にあたる 1998 年には漢字文献情報処理研究会が設立され、次の世代への基礎を成していくことになるが、ここでは漢字文献ということで、仏教や中国の哲学・思想テキストのデータ等が様々に議論されていた。

その一方で、人文学における研究基盤の構築のための要として、特に欧州ではその後着々と普及していった TEI だが、日本ではこの後しばらくの間、おそらくは上述の理由により、あまり姿を見せなくなってしまう。

Digital Humanities の登場

2000 年代初頭まで、Humanities Computing と呼ばれていた人文情報学は、2004 年の A Companion to Digital Humanities²⁶ の刊行とともに Digital Humanities (以下、DH)²⁷ と呼ばれるようになる。この名称は、編集者による提案であったとのことだが、それまでの Computing という名称と比べた時に、人文学中心の研究領域であることが明らかである DH という名称は、当時勃興しつつあった人文情報学の新しく大きな流れを象徴する変化として、その後一気に広まり、定着することとなった。その流れとは、Web が普及し、世界規模のインタラクティブな情報ネットワークを研究基盤とする人文学を展開する素地が整ったことである。この時期には欧米の人文情報学系学会の連合組織として ADHO (Alliance of Digital Humanities Organizations)²⁸ が設立され 2006 年には第一回の会議を開催するなど、DH の名を冠する場が急速に広がっていき、各地で DH のカリキュラムや研究センターが整備され、多くの DH 研究プロジェクトが助成金を得て展開されるようになる。2008 年には米国政府の人文学向け助成基金である NEH において Office of Digital Humanities が設立され、DH 専門の助成金が本格的に設定されるようになり、欧州でも DH 向けの研究助成金が広く配分されるようになり、さらに二国間²⁹ や複数国間³⁰ での共同助成金も開始されるなど、研究助成の側からも DH が強く支持されるようになり、これを踏まえた人文学研究基盤の整備も着実に進められることとなる。ここでも上述の TEI をはじめとして、それまでに策定されてきた様々な文化資料向けの規格

が、そうした研究助成の期待に対応できる堅実な受け皿としての力を発揮することになる。そのような中で、哲学・思想研究に資する大きな出来事としては、Transcribe Bentham プロジェクトが挙げられる³¹。Jeremy Bentham の草稿を翻刻（文字起し）して出版するプロジェクトが DH の力を借りてクラウドソーシングによる翻刻を実施したことにより、それまでとは比較にならない大量のテキストデータを短期間で作成することができたのであった。DH におけるこの種のプロジェクトとしては初期の大きな成功であり、その結果、様々な研究助成金を得て³² さらに発展することとなった。世界中で多くの研究者が専門的に取り組んでいる Bentham ならではの成功かもしれないが、Bentham 研究者の枠を超えてその存在自体を広く知らしめたという点でも貢献したと言えるだろう。ここでもやはり、TEI ガイドラインをはじめとするいくつかの文化資料向けの規格に準拠することで、横断性や持続可能性の高いデータの作成を実現していたのである。

一方、このような流れの中で増大し続ける分量と言語的多様性のために精読することが難しいテキスト群を研究するための方法論として、遠読 (Distant reading)³³ が選択肢の一つとされるようになった。当初は世界文学研究において提唱されたものであり³⁴、様々なコンピュータのツールを用いた統計分析を通じてテキストの特徴を把握することによってテキストに内在するものを理解しようとする研究動向である。これは、Google Books 及びそのデータを大学図書館が主体となって扱うべく設立された大学図書館連合 HathiTrust において、蓄積された膨大なテキストデータを分析するための考え方として用いられる³⁵ など、ビッグデータ時代における DH の可能性を示すものとして受容されるようになっていった。

ローカルに向けた・ローカルからの展開

2010 年代に入ると、DH は国際的にもさらなる広がりをみせ、2012 年の日本 DH 学会 (JADH) の設立と ADHO への加盟をはじめとして、世界各地で DH 学会が設立され、ADHO へと加盟していくことになる。DH が志向するグローバル性が、中核としてのグローバルな組織の主導よりもむしろローカル同士の丁寧な連帯によって成立するという状況が徐々に共有されていく時期であったと言えるだろう。この時期は、日本以外にも、オーストラリア圏、フランス語圏、メキシコ、南アフリカ、台湾が加盟³⁶ し、さらに、欧州 DH 学会³⁷ 内でドイツ語圏やイタリア、チェコ、北欧、ロシアの地域学会が設立されそれぞれに学術大会を開催するなどの活動を開始している。

同様の状況は上述の TEI においても見られるようになる。そもそも上述のように、TEI ガイドラインは、コミュニティに集った研究者達のニーズに

あわせて策定されてきたものであり、欧米の研究者しかそこにいなければ欧米の資料が対象の中心にならざるを得ない。TEI ガイドラインには外字に関する項目もあり、そこではすでに日本語母語話者による貢献もあった。外字は、大量の漢字を扱わねばならず、Unicode に登録されていない漢字としての外字も様々に扱わねばならない漢字文化圏の宿痾であり、同時にそれが一つの特徴であるとも考えがちだが、欧州においても中世写本に記載された多くの特殊な文字の表現に苦勞してきた面があり³⁸、外字はむしろグローバルな課題の一つとして捉えられてきたとみることができる。

このような状況が続いてきた中で一つの画期となったのが、TEI 協会における 2016 年の東アジア／日本語分科会 (SIG EastAsian/Japanese) の設立である。それまで、TEI ガイドラインは、グローバルなテキスト性を踏まえた資料の構造化を目指してきていたと見ることができるが、この分科会の設立は、そのような方向性に対して、むしろ個々の言語文化圏に蓄積されてきた慣習、いわばテキスト伝統に対応するという姿勢を示すものだったとすることができる。これに続いて、さらに書記体系が異なるインドテキストの分科会も設立されることとなった。さらに、国際化ワーキンググループも設置され、多言語グロッサリーの開発やガイドライン翻訳の簡易な手法の確立と実践を目指して活動を行っている。このこともまさに、ローカル同士の丁寧な連帯によってグローバル性の実現を目指す典型的な例であると言えるだろう。このような流れを受けて、2020 年 2 月、TEI ガイドラインは、その version 4.2.0 で、ルビの構造をテキストの標準的な形式の一つとしてガイドラインに導入した。言うなれば、漢字文化圏という巨大なテキスト群を抱える地域の周縁においてそれを読解するために生み出されたローカルルールが、人文学の研究手法に関する国際的なコミュニティにおいて対処すべき一つの構造として受容されたということであり、換言すれば、欧米外の地域における多様なテキスト文化のローカルな慣習にも同等の価値を置くという国際コミュニティの姿勢を鮮明に示したものであるとも言える。このような方向性は、今後も着実に進められていくことだろう。

DH の理念的背景

ここまで、哲学・思想研究との関連に配慮しつつこれまでの人文情報学の経緯についてみてきた。次に、そのようにして現在に至った DH が依拠してきた理念的背景を確認してみよう。すなわち、人文学と呼ばれる分野は非常に幅広く、同じ資料を対象にしたとしても研究手法が異なればそれに応じた

異なる研究成果が導き出されることになる。資料に関しても、文献資料だけではなく様々な資料があり得る。それにも関わらず、人文情報学は、そこに共通の場を見出そうとすることで一つのコミュニティを形成してきた。これには様々な背景が存在するが、なかでもその理論的背景として、この20年ほどの間、国際的な議論の形成を基礎づけてきた二つの理念がある。それは Scholarly primitives と Methodological Commons である³⁹。ここでは、これらについて検討してみよう。

Scholarly primitives

必ずしも人文学に限られる話ではないが、少なくともテキスト資料を扱う人文学の研究の過程において、ある程度共通に見られる行為がある。それを意識的に抽出することで基本的要素としようとするのが Scholarly primitives である。これを人文情報学の文脈において提唱した John Unsworth は、Scholarly primitives として以下の7つを提示した⁴⁰。

1. 発見 (Discovering)
2. 注釈 (Annotating)
3. 比較 (Comparing)
4. 参照 (Referring)
5. サンプリング (Sampling)
6. 例示 (Illustrating)
7. 表現 (Representing)

これらは哲学・思想研究に適用してみることも十分に可能だろう。テキスト資料の中にこれまで発見されなかった概念や注目されてこなかった事実を発見し、それらに対してどういう発見であるかを注釈し、その発見を位置づけるための他の類似の概念や事実と比較し、さらに関連する事象を参照する。その過程においては、資料の中から代表的な事例をサンプリングして検討するが、この検討に際しても、比較し、注釈し、参照することになる。その過程や成果を例示し、最終的にはその成果を表現する。人文学における重要な要素は思索だが、思索の内容そのものは分野毎としても個々の研究者としても固有性が高い。しかしながら、このようにして共通してみられる要素に着目することで、その思索が生じる基礎となる要素や、思索の結果を表し出す要素としての7つの Scholarly primitives を見出し得るといっているのである。

Unsworth が提起した Scholarly primitives が直接に目指していたのは人文学全般のためのデジタルツールの開発のための基礎付けであり、ここに挙げられた要素はそこに資するものとして有用な概念である。したがって、哲

学・思想研究向けのツールを作ろうとする際に、これらの primitives に準拠したツールを開発することで、人文学における汎用のツールを作成することができるということになる。たとえば、筆者が技術開発を担当した SAT 大蔵経テキストデータベース 2018 年版（以下、SAT2018）を例にとれば、まず、発見（Discovering）を行うために全文検索システムでの様々な検索手法や、文章間の対訳のリンク等を提供している。注釈（Annotating）に関しては、基本的には利用者の側で行うことと位置づけているが、データベース側でも、版本・写本画像上の異体字や異文、対訳との関係について注釈できる機能等、内容を共有しやすい一部の情報については注釈を行えるようにしている。比較（Comparing）に関しては、全文検索結果を並べて比較、複数の本文テキストを並べて比較、版本・写本画像を並べて比較できるようにするなど、様々な比較機能を提供している。また、特に、画像に関する注釈・比較に関しては、国際的に広く使われているデファクト標準の IIIF (International Image Interoperability Framework) に準拠することで、他の様々なシステムとの高い互換性を実現している。参照（Referring）に関しては、Web 上の関連する仏典リソースとの接続を提示しており、サンプリング（Sampling）については、Word2Vec による関連語抽出機能がそれにあたるだろう。例示（Illustrating）は、検索結果や探索の結果から随時表示される各々のコンテンツがそれにあたり、表現（Representing）は、そのようにして表現されるコンテンツであると同時にそれらを含む一連の表示であるとも言えるだろう。哲学・思想研究のためのツールとしての SAT2018 は、このようにして Scholarly primitives と対応づけることが可能である。その本来の目的を考慮するならそれらの機能が様々な種類のデータや他の役割を持つツールに対して開かれているべきであり、その点についても今のところ部分的には主に上述の IIIF を採用することで実現されている。

Primitives の議論そのものは、その後、より広い文脈に展開されていく。すなわち、これを踏まえつつ、学術研究全般における研究活動支援を目的とした研究図書館の文脈から研究活動における情報の扱いについての質的分析が行われ、より大きなプロセスの基礎や始まりにあるものとして primitives の意味が強調されることになる⁴¹。ここでは、研究活動は、Searching, Collecting, Reading, Writing, Collaborating, Cross-cutting Primitives 6 つに分類されており、primitives は、これらの具体的な活動に即してさらに詳細に分割されている。たとえば、「検索 Searching」を研究情報の「活動」として解釈しつつ、検索のプロセスを構成する活動の要素としての「直接検索 Direct searching」「関連付け Chaining」「閲覧 Browsing」「精査 Probing」「アクセス Accessing」を primitives とみなすということになる。この議論は、

デジタル人文学のためのインフラストラクチャの構築においてさらに検討され、DARIAH scholarly primitivesとして再定置される。ここで挙げられる primitives は、discovering, collecting, comparing, delivering, collaborating の5種類であり、さらにそれぞれに対して活動 (activities) が割り当てられる。たとえば、発見 discovering に対しては、Searching、Questioning、Browsing という活動が割り当てられるのである。この定義の違いは、研究活動を支援することを前提としたものか、デジタルインフラストラクチャを想定したものか、という差異に起因する。すなわち、primitivesとして分析しようとする場合、その概念の想定使用者の状況に応じて様々な形があり得ることになる。Unsworthの提示した Scholarly primitivesの場合には、人文学全般で利用可能なツールを作るための準備段階でありそれを目指した分析結果でもあるという位置づけであり、それらのいずれとも異なる立場からの primitivesということになる。このようにみえてみると、SAT2018を通じた確認を踏まえつつ、哲学・思想研究においてもそれぞれの場面に応じた primitives を措定してみることもまた、人文情報学の可能性を広げ、ひいては哲学・思想研究そのものをより豊かなものにするにもつながるかもしれない。

Methodological Commons

Digital Humanitiesという言葉が登場する直前となる2002年、その前身となる Humanities Computingの代表的な研究者であったキングスカレッジ・ロンドンのWillard McCartyとHarold Shortは、人文情報学における方法論の共有地としてのMethodological Commonsを提唱した⁴²。これは、人文学の様々な手法と分野を横断的に接続する「雲」の絵でよく知られるものであり、人文情報学において国際的に広く指示される理念的背景となった。Scholarly primitivesが抽象性を追求するものであるとするなら、Methodological Commonsはむしろ具体性にひたすら着目しようとする立場であるとみることができる。すなわち、人文学で、そこに含まれるそれぞれの分野において、様々な研究手法が提供されている。それらを共通の議論の場に引き出し、共有の知としてしまおうというのである。

Methodological Commonsにおいて手法として挙げられているのは、たとえば、哲学であれば認識論や存在論、心の哲学、哲学史や科学哲学、文学であれば想像力やレトリックとデザイン、歴史学なら歴史編纂や民族誌、科学史と技術、言語学ならコーパス言語学やコンピュータ言語学、社会学からは観察法や知識の構築などであり、一方で、情報科学からはプログラミングとデザイン、インターフェイス、コンピュータ言語学が挙げられ、図書館情報

学からは分類体系やメタデータ、デジタル図書館研究等が挙げられている。こうした様々な手法は、いずれもなんらかの形でデジタル技術と結びつけることが可能であり、その段階では各分野の様々な手法を横につなげて再検討することが可能である。たとえば、テキストの分析をコンピュータで扱おうとするなら、いくつかの手法で可能であり、その手法を共有することは有用である。それが方法論の共有地、ということになる。同様に、何かを分類しようとする手法であれば、分類に関わるデジタル技術を共通して利用可能であり、その適切な扱い方を検討し共有することもまた、方法論の共有地の役割である。そうした共有地を踏まえることで様々な研究結果がよりよい形で産み出されていくとしたら、その役割は現在重要であるというだけでなく、情報技術の長足の進歩を考慮するなら、今後はより一層重要になっていくことだろう。一方、この役割だけに着目するなら、単に便利な手法を共有する場という位置づけというだけのことになってしまうが、実際のプロセスにおいては、さらに意義のある事態が生じ得る。複数の分野が似たような手法において同じ技術を共有しようとするとき、そこには必ず、各手法における技術利用の際の微妙な差異が生じる。同じテキスト資料を扱っていたとしても、コーパス言語学では、まずはテキストの本文を確定した上で個々の単語の品詞情報まで利用しようとするのが多く、一方で、文献史的な観点からの研究であれば、テキストの本文の確定自体が研究となったり、そもそも本文を一つに確定しないことが重要であったりする場合もあり、そして、個々の単語の品詞情報までは明示的に課題としないという場合もある。同じテキストデータを扱い、同じようにコンピュータで検索したり統計処理をしたとしても、方法論としてはかなり異なるものになる。しかしながら、同じようなテキスト処理技術を用い、同じ場で議論する機会が提供されることになるのだとしたら、そこでは、相互に相手の方法論の意味を確認しあうというプロセスが生じることになり、相互に、自らが重視しない手法について改めて説明する必要が生じることもあるだろう。さらに、同じようなテキスト処理技術を用いているのであれば、これまで重視してこなかった手法を比較的容易に自らの方法論の中に取り込んでみるということもあり得るだろう。そのようにして、自らの方法論への反省の機会をもたらすことも可能であるという点が、方法論の共有地たる Methodological Commons の意義であり、それを現実の場としているのが、デジタル・ヒューマニティーズに関する各種研究集会や論文誌等の各種刊行物であり、コミュニティ Web サイトやブログ等も含めた緩やかなつながりがさらに補強している、ということになるだろう。

哲学・思想研究においても Methodological Commons で展開される互酬

性に基づく方法論的議論の場に参画していくことができたなら、デジタル時代における新たな豊かさを得られると同時に、他分野とのより深く新しい連携の仕方を通じた、研究内容のみならず組織的にも有機的に連関する学問の形成へとつながっていくことだろう。

そして、このような共有地における情報の交換に際して重要な拠り所となり得るのが、前節で論じた Scholarly primitives である。上述のように、primitives は状況に応じて異なるものであり、ここで求められるのは方法論の共有地における primitives ということになる。本稿ではそれが何であるかということまでは踏み込まないが、そこで展開される様々な活動において明示的にも暗黙的にも様々な形で埋め込まれており、それを踏まえた共有が様々ななされているはずである。その説明は今後の課題としたい。

終わりに

本稿では、哲学・思想研究と人文情報学との関わりについて、人文情報学の流れと基本を踏まえつつ検討した。人文情報学はデジタルという新しい要素を持ち込むものであり、1980年代より哲学・思想研究においても様々な形でのアプローチが行われてきた。しかしながら、それは単にある程度の利便性を高めるというだけでなく、研究という営みのために依拠する情報の流通の仕方を変革するとともに、そうした一連の変革を踏まえて、これまでの営みを振り返り、方法論的反省を通じてそれをさらに深める契機をもたらすものでもある。そのような状況の変化への対応には個々の研究者に応じて多様なあり方がある得べきだが、方法の分析とそれを共有する場を経由することで、研究者個人の取り組みを人文学全体に還元できるようなものへと高めていくことができる。人文情報学が哲学・思想研究にもたらし得る貴重な貢献があるとしたら、まさにその点にある。

謝 辞

本研究は JSPS 科研費 JP19H00516, JP19H00526, JP20H05830, JP20K20411, JP19H01241, JP20K20325 の助成を受けたものです。

注

- 1 下田正弘. 仏教とエクリチュール: 大乘経典の起源と形成. 東京大学出版会, 2020年.
- 2 Ong, Walter J. 声の文化と文字の文化. 翻訳者: 桜井直文ほか, 藤原書店, 1991年.
- 3 Eisenstein Elizabeth L., ほか. 印刷革命. みすず書房, 1987年.
- 4 その概要については、森岡 2019 がうまくまとめているので参照されたい。森岡正

- 博. 「人工知能と現代哲学」. 哲学, vol. 2019, no. 70, 2019年, pp. 51-68, doi:10.11439/philosophy.2019.51.
- 5 Kurzweil Ray. The singularity is near : when humans transcend biology. Viking, 2005年.
 - 6 宗教関連学会でも人工知能をテーマとする研究集が見られるようになってきている。近年の議論としては、以下の論文を参考として挙げておく。師茂樹. 「AI・仏性・倫理」. 未来哲学, no. 1, ぶねうま舎, 2020年, pp. 169-87.
 - 7 デジタル学術研究基盤の構築に関する具体的な手法や考え方については、永崎研宣『日本の文化をデジタル世界に伝える』樹村房, 2019. を参照されたい。
 - 8 Ide, Nancy, ほか. 「TEI:それはどこからきたのか。そして、なぜ、今もなおここにあるのか?」. デジタル・ヒューマニティーズ, vol. 1, 2018年, pp. 3-28, doi:10.24576/jadh.1.0_3.
 - 9 これに関しては枚挙に暇がないが、たとえば、土井裕人. 「視覚化と可触化による宗教学の教材作成とその応用に向けて」. 人文科学とコンピュータシンポジウム論文集, 2017年, pp. 231-38.
 - 10 Moretti Franco. 秋草俊一郎他共訳. 遠読:「世界文学システム」への挑戦. みすず書房, 2016年.
 - 11 A. Busa, Roberto. 「Foreword: Perspectives on the Digital Humanities」. ed. Susan Schreibman, Ray Siemens, John Unsworth. A Companion to Digital Humanities, Blackwell, 2004年.
 - 12 <https://eadh.org/about> (2021-01-10 確認)
 - 13 <https://ach.org/about-ach/history/> (2021-01-10 確認)
 - 14 <https://tei-c.org/> (2021-01-10 確認)
 - 15 この研究会及び後継のJACHについては、学位評価・授与機構の土屋俊先生に当時の資料をいただくことができたため、それに基づいていた記述となっている。
 - 16 内田保廣. 「情報処理語学文学研究会」のテキスト・アーカイブス」. 情報知識学会誌, vol. 3, no. 1, 1993年12月, pp. 45-51.
 - 17 豊島正之. 「TEIからみたSGMLのはなし」. 情報処理語学文学研究会報, no. 12, 1992年12月, <https://www.joao-roiz.jp/mtoyo/TEI/JALLC-12-TEI.pdf>.
 - 18 「TEI-P3について」. 情報処理語学文学研究会報, no. 15, 1994年7月, <https://www.joao-roiz.jp/mtoyo/TEI/JALLC-TEIP3.pdf>.
 - 19 坂井昭宏. 「コンピュータのなかの古典—テキスト・データベース研究会 (JACH) の研究活動—」. 情報知識学会ニューズレター, no. 1, 1988年9月, pp. 11-13.
 - 20 長瀬真理. 「TEIの活動と今後の展望」. 情報知識学会ニューズレター, no. 10, 1991年10月, pp. 9-10.
 - 21 ヘーゲル・テキストデータベース <https://www.unii.ac.jp/iori/jheg1.html> (2021-01-10 確認)
 - 22 浜渦辰二. 「峠を越えたフッサル・データベース: インターネット時代のマルチリンガル・テキストのために」. 人文論集, vol. 48, no. 1, 1997年7月, pp. A1-29.
 - 23 小野 基・小田淳一・高島 淳 「ダルマキールティ梵文テキスト KWIC 索引」, 東京外国語大学アジア・アフリカ言語文化研究所 1996.3.15. ISBN 4-87297-444-1
 - 24 中谷英明, 江島恵教. 「パリー三蔵データベースの構築と仏典研究」. パリー学仏教文化学, vol. 8, 1995年, pp. 123-47, doi:10.20769/jpbs.8.0_123.
 - 25 下田正弘, 永崎研宣. デジタル学術空間の作り方: 仏教学から提起する次世代人文学のモデル. 文学通信, 2019年.
 - 26 A Companion to Digital Humanities, ed. Susan Schreibman, Ray Siemens, John Unsworth. Oxford: Blackwell, 2004.
 - 27 人文情報学という言葉はDHを包摂しているが、2004年以降のDHを特に指し示す場

- 合には「DH」を用いる。
- 28 <https://adho.org/> (2021-01-10 確認)
- 29 NEH/DFG Bilateral Digital Humanities Program <https://www.federalgrants.com/NEH-DFG-Bilateral-Digital-Humanities-Program-36894.html> (2021-01-10 確認)
- 30 Digging into Data Challenge <https://diggingintodata.org/> (2021-01-10 確認)
- 31 Terras, Melissa. Present, not voting: Digital Humanities in the Panopticon: closing plenary speech, *Digital Humanities 2010. Literary and Linguistic Computing*, vol. 26, no. 3, 2011 年, pp. 257-69, doi:10.1093/lc/fqr016. この論文の元になった講演録の児玉聡氏による日本語訳は以下の URL で参照できる。 https://www.dhii.jp/dh/dh2010/DH2010_Plenary_trans_by_kodama.html (2021-01-10 確認)
- 32 Funding <https://blogs.ucl.ac.uk/transcribe-bentham/funding/> (2021-01-10 確認)
- 33 Moretti Franco. 秋草俊一郎他共訳. 遠読:「世界文学システム」への挑戦. みすず書房, 2016 年.
- 34 なお、コンピュータ利用の以前から遠読と呼びうる研究手法が存在したという解釈もある。 Cf. Underwood, Ted. 「A Genealogy of Distant Reading」. *Digital Humanities Quarterly*, vol. 011, no. 2, 2017 年 6 月.
- 35 こちらのリストを参照されたい。 <https://www.hathitrust.org/usage-examples-of-hathitrust-datasets> (2021-01-10 確認)
- 36 <https://adho.org/> (2021-01-10 確認)
- 37 <https://eadh.org/> (2021-01-10 確認)
- 38 <https://mufi.info/> (2021-01-10 確認)
- 39 Anderson, Sheila, et al. Methodological commons: arts and humanities e-Science fundamentals. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, vol. 368, no. 1925, The Royal Society, 2010 年, pp. 3779-96.
- 40 Unsworth, John. Scholarly Primitives: what methods do humanities researchers have in common, and how might our tools reflect this? <https://johnunsworth.name/Kings.5-00/primitives.html>. *Humanities Computing: formal methods, experimental practice*, King's College, London, 2000 年 5 月 13 日.
- 41 Carole L., Palmer, ほか. Scholarly information practices in the online environment: themes from the literature and implications for library service development. Technical report. OCLC Research, 2009 年 1 月, <http://www.oclc.org/research/publications/library/2009/2009-02.pdf>.
- 42 Willard, McCarty, Short Harold. 「Methodologies」. EADH - The European Association for Digital Humanities, 2002 年 4 月, <https://eadh.org/methodologies>.

(ながさき・きよのり 一般財団法人人文情報学研究所主席研究員)