# 歴史研究とデジタル・ヒューマニティーズの相関
# —東京大学大学院横断型教育プログラムを例として—

小風尚樹

## 0. はじめに

本稿は、筆者が「東京大学大学院横断型教育プログラム：デジタル・ヒューマニティーズ（以下、DH と略）」における講義・研究指導を経て[1]、2016 年 4 月にアメリカはボストンにて開催された国際ワークショップに招待を受け、研究発表を行うに至った経緯について、『人文情報学月報』第 59 号に寄稿した著者によるイベントレポートよりも[2]、個人的な問題関心・経験に即して綴るものである。

叙述の重点は、東京大学における DH 教育と筆者の問題関心の相互作用に置かれる。その理由は、日本における DH の普及が、様々な情報発信源が整備されつつも、大学院生や若手研究者への教育に関して言えば、一部に認知されども、十分に市民権を得ているとは言えず、今後の大学院教育に展開する際の検討モデルが現状乏しいと思われるからである。

加えて、本稿が、広く人文社会系の大学院生あるいは若手研究者に対し、研究にまつわる情報検索・収集に関する技術環境が刻一刻と変化している現状において、研究資源に対する情報リテラシーを涵養する必要性を認識させ、世界的な学術動向の中に自らの位置づけを措定するための一つの指針を提供することを願ってやまない[3]。

本論の構成は、以下の通りである。

1. わたしにとっての西洋史学研究と DH——言説研究と史料校訂に関する気づき
2. 史料を編む者と繙く者のはざまで——TEI 拡張スキーマに関する論文翻訳
3. DH 実践における諸サポート——アメリカでの研究発表

## 1. わたしにとっての西洋史学研究と DH——言説研究と史料校訂に関する気づき

### 1.1 歴史学と情報処理技術の共存可能な分析領域

修士課程進学時、もともと筆者は DH のことを全く知らなかった。

専攻は西洋史学で、卒業論文では日本開国期におけるイギリス海軍政策を扱った[4]。幸いなことに、近代イギリス史は、オンライン上の研究資源も非常に充実しており、議会議事録や議会提出文書、新聞や伝記辞典、英語辞書など、学部生でも豊富な研究資源に簡単にアクセスすることが可能である。

その意味では、学部生時代からオンラインデータベースの恩恵を受けていたため、大学院授業一覧を眺めていた際、「人文情報学概論」の文字に目を留めたのは偶然ではない。もちろん、かつて西洋史学の教授を務めておられた近藤和彦先生が、折に触れて情報技術に対する習熟の必要性を説かれ、『イギリス史研究入門』にもオンラインデータベースに関する情報に積極的に言及し[5]、紙面を比較的多く割かれていたことも大きな要因であったことは疑うべくもない[6]。

さて、筆者が最初に受講した「人文情報学概論Ⅰ」では、DH に関する学術動向を知るための情報源や、DH やウェブそのものの歴史、人文学史資料のデジタル化に関する国際的デファクト標準である TEI (the Text Encoding Initiative)[7]、テキスト計量分析や自然言語処理のツールなど、DH の方法論を支えるような基本情報に加えて、人文学が情報技術との邂逅によっていかに変容を迫られるか、あるいはむしろ伝統的な人文学的方法論への習熟

1 本プログラムの概要や科目一覧に関しては、以下を参照のこと。http://dh.iii.u-tokyo.ac.jp/

2 拙稿「『財務記録史料デジタル化の方法論をめぐる国際ワークショップ』参加報告」『人文情報学月報』第 59 号【後編】(http://www.dhii.jp/DHM/dhm59-2)、2016 年 6 月。

3 2015 年 9 月、デジタル手法を用いた歴史研究を評価するためのガイドラインが、アメリカ歴史学協会によって提唱されたように、歴史研究の収める射程が拡張していることから目を背けるわけにはいかないだろう（https://www.historians.org/teaching-and-learning/digital-history-resources/evaluation-of-digital-scholarship-in-history/guidelines-for-the-evaluation-of-digital-scholarship-in-history)。なお、同ガイドラインは、日本における DH 研究成果に対する学術的評価をめぐる議論を活性化させるという目的で、日本語訳が行われた。菊池信彦・小風尚樹・師茂樹・後藤真・永崎研宣訳「歴史学におけるデジタル研究を評価するためのガイドライン」『東京大学学術機関リポジトリ』http://repository.dl.itc.u-tokyo.ac.jp/dspace/handle/2261/59142, 2016 年 2 月。

4 拙稿「19 世紀中葉イギリスの東アジア戦略における日本の位置づけ——イギリス海軍司令長官スターリングの北東アジア観と函館港」『クリオ』第 29 号、2015 年、44-58 頁。

5 近藤和彦編『イギリス史研究入門』山川出版社、2010 年。

6 なお、ヨーロッパ史の各『研究入門』におけるオンラインデータベースに関する情報の比較については、菊池信彦「西洋史 DH の動向とレビュー——西洋史学はウェブ情報をどのように位置づけているのか　『研究入門』を題材に」『人文情報学月報』第 46 号【前編】(http://www.dhii.jp/DHM/dhm46-1)、2015 年 5 月、を参照のこと。

7 TEI については、本稿 1.2. にて詳しく述べる。

によってこそ見えてくる情報技術との共存可能性などについて、充実した講義が展開された。

もっとも、筆者自身が興味を引かれたのは、講義も全体の 3 分の 2 を過ぎたあたり、テキスト計量分析ツール (Voyant Tools) を紹介されたことがきっかけであった[8]。

折しも、西洋史学の指導教員勝田俊輔先生の大学院ゼミにおいて、democracy という語が、「長い 19 世紀」の中でいかにその含意するニュアンスを変遷させていったのか、という一種の言説研究を講読テキストとして扱っていた頃であった。フランス革命を契機とし、むしろ暴力的な民衆政治をイメージさせる語として扱われていた democracy が、19 世紀ブリテンにおける選挙法改正の動きの中で、民衆の政治参加の割合が漸進的に増加していくにつれて、ポジティブな含意を伴って用いられるようになったとする同書の記述は明快で、論拠とする新聞やパンフレットといった史料を豊富に利用する手法に舌を巻く思いであった[9]。

しかしながら、言説分析のための史料の数が多くなればなるほど、分析に際して、恣意性を排除し、網羅性を担保するのは難しくなるのではないか、とも同時に感じていた。

そうした折、「人文情報学概論」の中で、テキスト計量分析ツールや MIMA サーチについての紹介が行われ[10]、光が見えたような思いを抱いたのである。具体的には、分析対象となるテキスト群（テキストコーパス）に対し、あらかじめ仮説を持ってテキスト分析に臨むというよりは、ある単語の使用頻度や共に使われやすい語（共起語）について、人手ではなくコンピュータに分析を行わせることで、見落としや恣意的な分析を防ぐ可能性があると感じたのである[11]。つまり、コンピュータによる定量的な分析が、歴史学の方法論を説得的に裏付ける可能性を感じたということである[12]。

ごく簡単に言っても、分析対象となるテキストコーパスの特徴を、精読する前に概観しておくこともできる。例えば、すでにオンラインテキストデータとして活用がしやすくなっているイギリス議会議事録に着目してみたい[13]。1840 年 4 月 8 日、庶民院にて、第一次アヘン戦争での中国介入をめぐって、ウィリアム・グラッドストンがかの有名な反対演説を行った議事録を例に取ろう[14]。

先に紹介した Voyant Tools を用い、上記議事録のテキストデータを読み込ませた結果をいくつか表示する[15]。

**表 1　テキストコーパス中の頻出語一覧**

まず、表 1 は、どのような単語が頻出しているのかを把握するための基本的な機能である。のべ 36000 語以上の単語が使われているようだ。続いて、この頻出語一覧の中から、最もよく使われている単語 trade に着目する。テキスト計量分析を行

8 http://voyant-tools.org/ なお、同ツールは、2016 年に大型アップデートが行われ、日本語などのマルチバイト言語のテキスト分析にも対応するようになった。重要なのは、スペースで単語が区切られていない自然言語を、形態素解析を行うことで品詞分解ができるようになり、その上、わかりやすいユーザ・インタフェースで多種多様な解析ツールが提供されていることである。このアップデートを機に、筆者を含むチーム（人文情報学研究所の永崎研宣氏を中心に、佐藤正尚・杉浦清人・鈴木親彦・王一凡諸氏）が、日本語メニュー対応のための翻訳作業を完成させた。

9 Arthur Burns and Joanna Innes, eds., *Rethinking the Age of Reform——Britain 1780-1850*, Cambridge University Press: Cambridge, New York, 2003.

10 MIMA サーチは、東京大学大学院工学系研究科の美馬秀樹准教授が中心となって開発を行った、自然言語処理技術に基づく情報検索・可視化システムである。大規模テキストデータを対象とし、データベースに読み込ませる必要があるため、Voyant Tools のように研究者個人がテキスト計量分析を試験的に行いたいような場合には不向きだが、東京大学の DH 教育プログラムにおけるコア科目である「知の構造化論」においても、MIMA サーチを活用したテキストコーパスの構造的把握が扱われている。「東京大学授業カタログ (http://catalog.he.u-tokyo.ac.jp/)」や、サステイナビリティ・サイエンスの学術情報を扱う「Network of Networks (http://nns-u.org/)」など、MIMA サーチの活用事例は多岐にわたるが、国立国会図書館も書誌情報検索・可視化システムとして導入しており、そこには利用マニュアルも掲載されている (http://lab.ndl.go.jp/ut/)。

11 質的内容分析と比較した場合の、量的分析の有用性や、テキスト計量分析ツールのマニュアルとしては、以下を参照のこと。樋口耕一『社会調査のための計量テキスト分析——内容分析の継承と発展を目指して』ナカニシヤ出版、2014 年。

12 ビッグデータの活用によって、歴史における「長期持続」の様相を提示し得る可能性も注目されているように (Jo Guldi and David Armitage, *The History Manifest*, Cambridge University Press: Cambridge, 2014, pp. 88-116)、人文学的な定性的分析が、科学的な定量的分析と共存し得ないと考えるのは早計だろう（徃住彰文監修・村井源編『量から質に迫る——人間の複雑な感性をいかに「計る」か』新曜社、2014 年）。日本の西洋史学研究者の間にも、テキスト計量分析への関心が現れている。例として以下を参照のこと。楠義彦「前近代歴史資料の計量的分析——Visitation Articles を事例として（特集 デジタルメディアと歴史学）」『ヨーロッパ文化史研究』第 14 号、2013 年 3 月、29-53 頁。

13 HANSARD 1803-2005 (http://hansard.millbanksystems.com/)

14 グラッドストンの演説でよく知られるのは、「なるほど中国人には、馬鹿げた大言壮語と高慢なプライド、そして度を超す罪があることは疑いようもありません。しかしながら、正義は、異教徒であり半文明な野蛮人［中国人：筆者註］側に存すると考えるのです」の箇所であろう。
http://hansard.millbanksystems.com/commons/1840/apr/08/war-with-china-adjourned-debate#S3V0053P0_18400408_HOC_6

15 なお、同ツールはアップデートにより、複数テキストのファイルを解析することができるようになっているが、今回は一つのテキストファイルを扱う。

う場合、予備知識があると、たとえ頻出語でなくとも注目に値する単語などが事前にわかることがあるが、基本的には、定量的に注目に値するものから探索を進めていくのが、恣意性の排除という点からして妥当である。

**表 2　頻出語同士の共起情報ネットワーク**

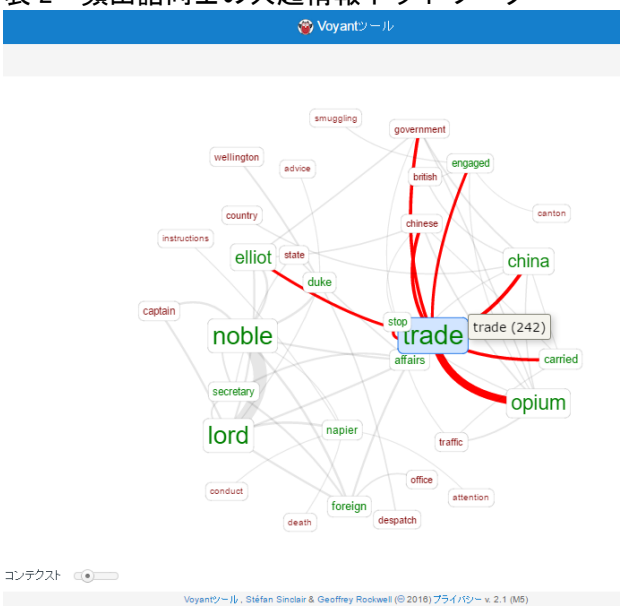
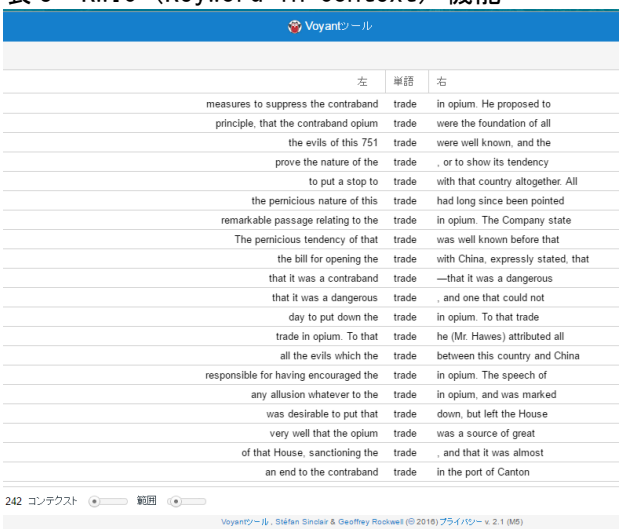



表 2 は、頻出語がどのような単語と共に用いられるかを可視化したものであり、同ツールにおいては、単語同士の結びつきが強いほど、結ばれる線（エッジ）が太く表示されるようになっている。アヘン貿易およびその廃止といった文脈のほか、第一次アヘン戦争当時のイギリス海軍士官で、正副全権大使であった George Elliot と Charles Elliot のいずれか、あるいは双方に言及する形で[16]、trade が用いられていることが見受けられる。続いて、より詳細に、trade がどのような文脈で用いられているのかを確認する必要がある。

**表 3　KWIC（Keyword in Context）機能**



Voyantツール

| 左 | 単語 | 右 |
|---|---|---|
| measures to suppress the contraband | trade | in opium. He proposed to |
| principle, that the contraband opium | trade | were the foundation of all |
| the evils of this 751 | trade | were well known, and the |
| prove the nature of the | trade | , or to show its tendency |
| to put a stop to | trade | with that country altogether. All |
| the pernicious nature of this | trade | had long since been pointed |
| remarkable passage relating to the | trade | in opium. The Company state |
| The pernicious tendency of that | trade | was well known before that |
| the bill for opening the | trade | with China, expressly stated, that |
| that it was a contraband | trade | —that it was a dangerous |
| that it was a dangerous | trade | , and one that could not |
| day to put down the | trade | in opium. To that trade |
| trade in opium. To that | trade | he (Mr. Hawes) attributed all |
| all the evils which the | trade | between this country and China |
| responsible for having encouraged the | trade | in opium. The speech of |
| any allusion whatever to the | trade | in opium, and was marked |
| was desirable to put that | trade | down, but left the House |
| very well that the opium | trade | was a source of great |
| of that House, sanctioning the | trade | , and that it was almost |
| an end to the contraband | trade | in the port of Canton |

242 コンテクスト　範囲

Voyantツール , Stéfan Sinclair & Geoffrey Rockwell (© 2016) プライバシー v. 2.1 (M5)

表 3 は、ある単語（この場合は trade）が、どのような文脈で用いられているのかを表示したものである。ここでは、アヘン密貿易の危険性やその廃止に言及する用例が多いことがわかる。

こうしたテキスト解析は、人間が精読すれば内容をより深く理解することが可能だろうが、それでも多くのテキストを対象とする場合や、精読のための準備とする場合にも有効である。

しかしながら、イギリス議会議事録のように、テキストデータとして用いることのできるデジタル化された史料というのは、決して多くはない。こうしたテキスト計量分析のツールが充実する一方で、史料のデジタル化の歩みは現状どうなっているのだろうか。

### 1.2　保存媒体の相違が史料校訂・分析に及ぼす影響

そもそも、デジタル化された史料と聞いて、多くの研究者がまず思い浮かべるであろうものは、Internet Archive や gallica といったオンラインデータベースの中で[17]、紙媒体の史料をデジタル撮影した pdf ファイルや jpeg ファイルといった画像のデータベースではなかろうか。あるいはもう一歩進んで、Google Books や House of Commons Parliamentary Papers のように[18]、同じく画像データではあるけれども、OCR (Optical Character Recognition) 処理が施されていることから、画像認識技術でもって単語検索がかけられるといったものだろう。

16 横井勝彦『アジアの海の大英帝国――19 世紀海洋支配の構図』同文館、1988 年、63-73 頁。
17 Internet Archive (https://archive.org/); gallica (http://gallica.bnf.fr/).
18 House of Commons Parliamentary Papers (http://parlipapers.chadwyck.co.uk/home.do). Google Books については特に説明不要だと思われるが、Google Books のテキストデータを用いた Google Ngram Viewer (https://books.google.com/ngrams) は、単語の出現頻度の時代的推移を可視化することができる web アプリケーションであり、様々な活用法が考えられて興味深い（マニュアルは、以下を参照。https://books.google.com/ngrams/info#）。例えば、動詞の過去分詞形が変化した時期の特定、複数言語コーパスにおけるある人物に対する言及頻度の比較、歴史学の手法をめぐる潮流の比較（長谷川貴彦『現代歴史学への展望――言語論的転回を超えて』岩波書店、2016 年、157 頁）、などが挙げられよう。

これに対し、史料利用の可能性をさらに一歩進める役割を担っていると言ってよいのが、人文学史資料のデジタル化に関するデファクト標準 TEI である。すなわち TEI は、画像データでしかなかった史料に、意味を持たせながらテキストデータを付与するものなのである。基本的には、史料の画像データを対象に、翻刻（文字起こし）を行い、そのテキストデータをタグ付き言語で記述していくのである。TEI は、この記述のためのルールを柔軟に定めた事実上の国際標準なのだ[19]。

さらに言えば、国際標準たる TEI の強みというのは、対象となる史料の内容に即して、テキストデータをコンピュータ可読形式にタグ付け（マークアップ）するためのルールが設定され、そして何よりそのルールに準拠してマークアップを行う学術機関・プロジェクト・研究者の輪が国際的に広がっている、という事実にある[20]。

具体的なルールとして特に重要なのは、TEI では、マークアップ元となる史料の校訂情報や出版情報、マークアップ担当者の氏名やマークアップへの関わり方などについて、TEI ヘッダー内に情報を明記することができることである[21]。つまり、出典を明らかにすることができるのだ。このように、史料編集に関するメタデータを併記することこそが、論拠としての史料的信頼性を担保することにつながっていると考えられる。このことは、紙媒体の刊行史料を用いる際でも、採用する版によって解釈が変わり得ることに留意し、違う版と見比べることがあるように、史料校訂の責任が明記されていることで分析手続きの妥当性が担保されることと比較してみれば、明らかであろう。したがって、TEI に準拠してデジタル化された史料というのは、保存媒体をウェブに移行しただけであって、その史料的重要性を貶めることにはつながらないと考えられる。

もちろん、留意しなければならない点も存在する。それは、紙媒体とウェブ媒体の史料的性質の違いに起因する、研究者による情報収集行動に違いが生じることである。

これは、紙辞書と電子辞書の違いと本質的には同じ問題であると言えよう。紙辞書を引けば、調べたい単語の周りに掲載されている別の単語や用例にも目が留まることがある一方で、電子辞書で調べる場合は、調べたい情報をピンポイントで検索することが基本であるため、周辺情報は比較的見えにくい。

同様に、検索可能なウェブ媒体の史料にアクセスする場合、腰を据えて始まりから終わりまでざっと目を通すというよりも、キーワードの検索とその周辺情報の探索が主な情報収集行動になるだろうと思われる[22]。そもそも人間の脳は、高度な情報を処理する際、スクリーン上で見た場合と紙面上で見た場合とでは、前者の方が、理解度が低くなるということも指摘されているように[23]、史料の深い内容理解と歴史像の把握といった面では、紙媒体の史料を精読する有用性が優ると言えよう[24]。

しかしながら、そもそも、ウェブ媒体の史料に対して、紙媒体の史料に対するアプローチと同じものを採用する必要があるのだろうか。

すでに述べたように、テキスト計量分析が有効なのは、人間が処理しきれないほど膨大な量のテキストコーパスに対し、単語の出現頻度や共起情報の数え上げなどの単純な作業を、延々と正確に、かつ圧倒的に速く行うことである。他にも、テキストデータが入力されておらず、画像データしか存在していないような手稿文書の史料群に対しても、似た字形の画像を瞬時に切り出して表示する SMART-GS の技術開発が進んでいるように[25]、ウェブ媒体の史料を用いることの一番のメリットは、日進月歩の情報処理技術の恩恵に与るということであるため、あえてアナログな人間の分析と競合

---

19 TEI P5 ガイドラインについては、以下を参照のこと。http://www.tei-c.org/Guidelines/P5/ なお、史料のデジタル化の必要性については、紙媒体の史料の経年劣化や腐食を阻止するなど、史料保存の観点からしても重要であることは言うまでもない。マークアップについては、本稿の 2.2. にて詳しく後述する。

20 2016 年 6 月には、TEI 協会内に、日本語テキストのマークアップに関するガイドライン提言のための組織である東アジア／日本語分科会 (East Asian / Japanese SIG) が設置された。同分科会の第 1 回会合については、以下を参照のこと。http://21dzk.l.u-tokyo.ac.jp/DHI/index.php?teijpsig

21 ヘッダーとは、マークアップ言語の記述において、本文情報の前に配置され、人間が見るコンピュータの画面には現れない情報が記されている箇所のことである。html ではこのヘッダーに、文字コードの指定や関連するスタイルシートへのリンクなどが記されている。TEI/XML では、このヘッダーに、史料の校訂情報などのほか、史料中に登場する人物や場所の説明を記しておくことができる。

22 同様の指摘は、故二宮宏之氏が行っている。史料デジタル化に対する二宮氏の警鐘への応答も意識する形で、フランスにおける DH の取り組みを紹介したものとして、以下を参照のこと。長野壮一「デジタル歴史学の最新動向――フランス語圏におけるアーカイブ構築およびコミュニティ形成の事例紹介」『現代史研究』第 61 号、2015 年、39-48 頁。

23 スーザン・グリーンフィールド（広瀬静訳）『マインド・チェンジ――テクノロジーが脳を変質させる』KADOKAWA、2015 年、313-322 頁。原著は、Susan Greenfield, *Mind Change――How Digital Technologies are leaving their Mark on our Brains*, Random House: New York, 2015.

24 こうした指摘は折に触れてなされており、多くの歴史研究者が共感するところでもあろう。例として、以下を参照のこと。岡本隆司・吉澤誠一郎編『近代中国研究入門』東京大学出版会、2012 年、256-260 頁。

25 SMART-GS のプロジェクトについては、以下を参照のこと。https://osdn.jp/projects/smart-gs/ 同プロジェクトの一員である京都大学文学研究科の橋本雄太氏の以下の記事も参照のこと。橋本雄太「集合知で読む歴史史料――SMART-GS が実現するグループリーディング」『人文情報学月報』第 37 号【前編】(http://www.dhii.jp/DHM/DHM37_smartgs)、2014 年 8 月。

させる必要がない。分析対象や手法の性質に合わせて、紙媒体とウェブ媒体の史料を使い分けることが肝要だろう。

以上のように、「人文情報学概論 I」で着想を得た、情報学的知見が言説研究の手法を補強し得る可能性と、史料校訂に関する本質は、筆者にとって、西洋史学と DH が、相互に補完し合う可能性を十分に有すると認識させるものであった。

## 2. 史料を編む者と繙く者のはざまで

### 2.1 TEI 拡張スキーマに関する論文翻訳と学会発表

2014 年度の冬学期、続けて受講した「人文情報学概論 II」では、参加者が DH 関連学会での研究発表を行うことが講義の目標として設定され、そのためのサポートを担当教員が行うという、院生側からすると非常に恵まれたカリキュラムが組まれた。

筆者自身が最初に考えたのは、直前の夏季休暇に渡英して行った史料調査の記憶が新しかったこともあり、ロンドンの国立公文書館（The National Archives. 以下、TNA と略）所蔵のイギリス政府文書のデジタル化事業の提案を試みる、というプロジェクトであった。しかし、TNA の史料デジタル化事業は、プロジェクト提案側がデジタル化にかかる費用を全額負担しなければならないという決まりが存在し[26]、あえなく撤退せざるを得なかった。

そこで、担当教員の一人であった人文情報学研究所の永崎研宣先生から最初にいただいた課題が、財務記録史料のマークアップのための TEI 拡張スキーマに関する最新論文を翻訳することであった[27]。財務記録史料とは、広義の商取引に関する情報を残す史料群のことで、複式簿記や帳簿、領収書や日記から、貿易統計など非常に多岐にわたるものである。それゆえ、そのように議論の射程が広い財務記録史料のデジタル化に関する方法論が新しく提案されたということ自体、情報収集に値するものであるとのご指摘をいただき、翻訳に取り組むこととなった[28]。財務記録史料のマークアップ論文の翻訳を踏まえ、筆者は 2015 年 5 月に開催された、第 106 回人文科学とコンピュータ研究会にてポスター発表を行った[29]。

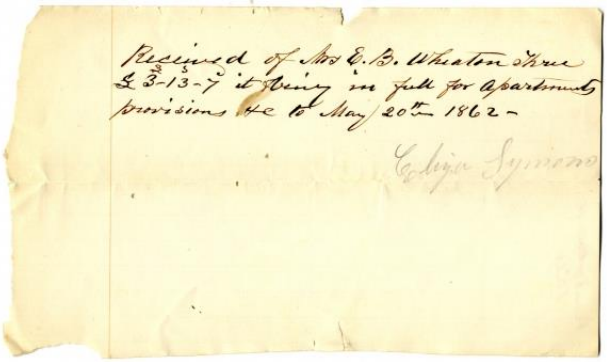
こうした一連の発表成果がきっかけとなり、幸運にも、翻訳論文の著者である Kathryn Tomasek 教授が主催した国際ワークショップでの研究発表の招待を後にいただくこととなる。

### 2.2 史料マークアップと史料読解

本節では、これまでのマークアップの経験から得た気づきを記しておきたい。結論を先取りすれば、史料への向き合い方は、マークアップを行う者と史料読解を行う者との間では、真逆の方向性を志向するものであると思われる。

そもそも、史料のマークアップと言っても、多くの歴史研究者にとって馴染みのないものと思われるので、以下に Tomasek 氏の論文から例を示す。

図 1　宿泊部屋と乗車券に関する領収書、1862 年 5 月 20 日、Wheaton Family Papers より[30]


Received of Mrs E.B. Wheaton three £3-13-7 it being in full for apartments provisions &c to May 20th 1862 –
Eliza Symons

例えば、上記のような手稿史料があったとしよう。この記述内容に即して、以下のように機械可読形式で記述することを、マークアップと呼ぶ。特に、下記のマークアップは、TEI のルールに則った形式の XML (eXtensive Markup Language) 文書なので、TEI/XML マークアップと呼ばれる。

```
<text facs="ebweur1862_009a.jpg">
   <body>
      <p>
         <handShift new="#pers_WCDH1"/>Received of Mrs E.B.
           Wheaton Three
           <measure type="currency"> £ 3
             <hi rend="superscript-above">£</hi>-13
             <hi rend="superscript-above">s</hi>-7
             <hi rend="superscript-above">d</hi>
           </measure> it <unclear>&amp;</unclear> in full
              for
           <measure commodity="room">apartments</measure>
           <measure commodity="board">provisions</measure>
            &amp;c to <date notAfter="1862-05-20">May20
              <hi rend="superscript">tt</hi> 1862.</date>
      </p>
      <signed>
         <handShift new="#pers_WCDH716"/>
            Eliza Symons
      </signed>
   </body>
</text>
```

26 http://www.nationalarchives.gov.uk/documents/opportunity-2010.pdf

27 Kathryn Tomasek and Syd Bauman, ‘Encoding Financial Records for Historical Research’, *Journal of the Text Encoding Initiative* [Online], Issue 6 | December 2013, Online since 26 September 2013, URL : http://jtei.revues.org/895

28 完成翻訳稿は、以下を参照のこと。拙訳「歴史研究のため財務記録史料マークアップ手法」『東京大学学術機関リポジトリ』http://repository.dl.itc.u-tokyo.ac.jp/dspace/handle/2261/56940, 2015 年 5 月。

29 拙稿「19 世紀イギリス政府文書における財政・統計関連史料のマークアップ例提示」『情報処理学会研究報告．人文科学とコンピュータ研究会報告』2015-CH-106(7), pp. 1-5, 2015 年 5 月 (http://ci.nii.ac.jp/naid/110009899549).

30 図 1 およびマークアップ文ともに、Tomasek and Bauman, *op. cit.*

このように、史料のマークアップとは、端的に言えば、人間が読んで理解できる内容を、コンピュータが理解できるような言語で書き直すことであると言ってもよい。そのため、人間にとってはむしろ煩雑に思えるだろうが、こうしたタグ付き言語でもって史料の意味内容を記述することで、様々なメリットが得られる。

例えば、史料に登場する人物名に名前の ID を付与して記述しておけば、その人物の ID で検索をかければ、時代や個々の場面によって使い分けられる様々な呼称の一覧を瞬時にリスト化することができる。

このようにタグ付けを行うのは、人物に対してだけではなく、組織・場所・年代・モノといったものから、Tomasek 氏たちの研究プロジェクト MEDEA (Modeling semantically Enriched Digital Edition of Accounts) で扱われているような、商取引の情報までもが対象となる。とにかく、史料に記載される内容を、その意味に即して、人間が理解しやすいタグで記述しつつ、機械可読性を維持することが、史料のマークアップである。

もちろん、ごく簡単に述べただけでも、マークアップ作業というのは膨大であることが伝わるだろう。加えて、意味を理解した上でタグ付けを行っていかなければならないので、タグ付けの自動化をするにも、モデルとなるマークアップ文書が作成される必要がある。

史料のマークアップを行うということは、web 媒体の史料を編集する上で本質的に重要な作業であると言いかえることもできる。その際、史料の内容を解釈する行為が含まれることは当然だが、それよりもむしろ、史料が語りかける情報を残さずすくい上げ、潜在的な史料の読み手に提供するものを少しでも多くすることが、マークアップの本質であり、「史料を編む」ということなのだろう。

これに対し、研究者などが史料を読解するという行為の本質は、「史料を繙く」ことであると言えよう。つまり、史料が語り掛ける情報を受け止め、その上で情報を取捨選択し、歴史像を再構成するための糸を引き出すことであると言い換えてもよい。

このように考えると、まさに史料マークアップと史料読解とは、史料の情報を余さず記述するか、情報の取捨選択を行うかという点において、まさに正反対の方向性を志向するものであると言えよう[31]。こうした正反対の性質を持つ作業に従事することで、研究活動に対してもたらされた意義について、個人的な所感を記しておきたい。

重要な点として挙げられるのは、分析対象となる史料の構造そのものに着目することから、歴史研究に必要な史料分析の能力を補うことにつながる点である。特に筆者が用いるような外交関連の政府公文書などは、イギリスや中国、日本の史料それぞれに特徴的なフォーマットが定まっており、定型表現も数多く存在する。史料読解の際には、まずこうした定型化された箇所から構造的に読み解き、内容理解に入っていく。その意味で、史料マークアップの際にどのようなタグを用いるかを考えることは、史料の型への意識を高めることにつながるという相互作用があると思われる。

## 3. DH 実践における諸サポート——アメリカでの研究発表

### 3.1 MEDEA ワークショップでの研究発表準備

すでに述べてきたように、2016 年 4 月の初頭、Tomasek 氏らが中心となっている研究プロジェクト MEDEA の第 2 回ワークショップが、ボストンの Wheaton カレッジにて開催された。筆者が DH で指導を仰ぐ永崎研宣氏は、Tomasek 氏と旧知の仲であったこともあり、件のワークショップへの若手研究者の招待の話を受け、筆者に参加を促していただいた。

ワークショップの内容については、筆者によるイベントレポートをご覧いただくとして[32]、本節では、ワークショップでの発表準備にまつわる諸サポートについて触れ、謝辞に代えさせていただきたい。

まず、東京大学人文社会系研究科インド哲学仏教学研究室／次世代人文学開発センター創成部門「人文情報学」の下田正弘教授は[33]、アメリカでの発表が決まった頃、ご多忙の中、研究の相談も兼ねてお時間を割いてくださり、DH の研究進捗だけでなく、筆者の西洋史学の修士論文についてもあたたかいコメントをくださった。下田先生からコメントを賜る機会は幾度かあったが、大きな枠組みの中に研究を位置づけることの重要性を常に説いてくださった。その際、先生のコメントは、一つの学問分野にとらわれず、人文学と情報学のあいだを自由に行き来するものであり、いつも目を開かせられる思いを抱かされてきた。

次に、同次世代人文学開発センターの A. Charles Muller 教授は[34]、発表申し込みの要旨や発表原稿の英文チェック、および英語でのプレゼンテーシ

31 こうした気づきを得るにあたっては、2016 年 7 月 3 日に開催された「仏教学新知識基盤の構築　次世代人文学の先進的モデルの提示（科学研究費基盤 S）」の研究会合で研究発表を行った際の討論も、非常に示唆的であった。

32 本稿註 2 を参照のこと。

33 http://www.l.u-tokyo.ac.jp/teacher/database/14.html

34 http://www.l.u-tokyo.ac.jp/teacher/database/1540.html

ョンの練習に多くのお時間を割いてくださった。そもそも普段から英語の添削をしていただく機会に恵まれることは難しい中で、ご自身も TEI のプロジェクトを進めてこられた Muller 先生に、専門的なご指摘も含めていただいたアドバイスは、実際のプレゼンテーションの場に、何も臆することなく臨むことのできた大きな支えとなった。

西洋史学の指導教員である勝田俊輔先生には、博士課程進学が決まって後、アメリカに出発する 2 週間ほど前に面談をしていただく機会があった。その場は、博士課程進学後の研究の方向性や計画についてご相談をすることが主だったが、筆者としては思いがけず、先生はアメリカでの DH の発表についてもコメントをくださった。それは、マークアップする史料の成立背景、史料内容の文脈を精確に理解することができているかということや、発表で予定していた、中国語と英語双方の史料の翻訳対応箇所のマークアップについて、そもそもなぜ翻訳が対応しているのか、どの程度まで対応しているのかなど、史料そのものに向き合う姿勢を改めて認識させてくださった。歴史研究者が史料と対峙するとき、それがマークアップであろうと史料読解であろうと、史料批判を行うという当たり前のことを、実践的に教えてくださったように思う。

最後に、人文情報学研究所の永崎研宣先生は、研究発表のお誘いに始まり、国内外を飛び回るかたわら頻繁に筆者との面談に時間を割いてくださった。TEI/XML マークアップ文書の作成にあたってのアドバイスだけでなく、作成した XML ファイルからデータを取得して、JavaScript 可視化ライブラリ d3.js に読み込ませる php プログラムまで作成していただいた。アメリカ滞在中は、ワークショップの合間を縫ってプログラムや発表原稿の最終チェックをしてくださり、Syd Bauman 氏や Øyvind Eide 氏など、多くの DH 業界の研究者との交流を持たせていただいた。さらに、ボストンでの MEDEA ワークショップに加えて、帰国前日には、UCLA の East Asian Library の日本研究専門司書 Tomoko Bialock 氏のご助力を賜り、永崎先生ご自身が携わるプロジェクトである大蔵経データベースの紹介と[35]、筆者の研究発表を行う私的なワークショップを開催していただく手配もしてくださった。

UCLA, East Asian Library の入り口から

同図書館は、日本をはじめとする東アジアに関する研究書を多く所蔵しているほか、DH に関するワークショップやミーティング、講義を行うスペースが非常に充実していたことが印象的であった。

DH 関連のプロジェクトを扱う際のラウンジ

35 SAT 大正新脩大藏經テキストデータベース (http://21dzk.l.u-tokyo.ac.jp/SAT/index.html)

このように、二度にわたるワークショップをはじめ、筆者自身にとって初めての渡米経験が非常に充実したものになったのは、永崎先生のご助力なくして考えられなかった。改めてここに感謝の意を表したい。

なお、本稿を執筆するにあたっては、DH に限らず、西洋史学研究に携わる院生諸氏からいただいた多くのアドバイスやコメントに助けられたことを記しておきたい。情報収集のツールや行動に大きな変化が訪れている昨今、アナログとデジタルを自由に行き来する若手研究者や大学院生にとって本稿が、人文学に吹き込む新しい風に思いを馳せ、研究の視野が広がるきっかけとなれば幸いである。