

タイトル：2020 年度教育セミナー（第 16 回）

日時：2020 年 9 月 17 日（木）～20 日（日）

オンライン開催

「最近の歴史情報学の動向について」

熊倉和歌子（AA 研）

本セミナーでは、最近の歴史情報学の動向として、TEI（Text Encoding Initiative）と IIIF（International Image Interoperability Framework）を中心に現在までの展開を解説しながら、現在講師が取り組む人名録のテキスト分析の研究などを紹介した。

昨今、人文情報学は、国内外において最も活気ある分野の一つとなり、人文学と情報学の架け橋となる分野として学術分野の中に着実に地歩を固めつつある。人文情報学のうち歴史的データを扱い歴史研究を推進する歴史情報学は、人文情報学推進の大きな原動力となっており、情報学の専門家や歴史学の専門家が連携して研究を進め、発展段階にある。そうした展開において、特に注目されているのが、TEI によるテキスト分析と、画像データを公開するさいの枠組みである IIIF である。TEI は、XML をベースとしたタグ付け指針である。デジタル化したテキストに対し、共通のタグをつけていくことにより、テキストの内容から特定のデータを抽出してアルゴリズム処理することを可能にする。一方、IIIF は、画像データ公開の共通の枠組みであり、IIIF ビューワの利用により、画像データに自由に注釈を書き込むことや、さまざまなウェブサイト上に公開されている画像データをユーザーのビューワ上に集めることが可能となる。TEI や IIIF に共通するのは、共通の規範・手法で分析や公開・共有を行うことにより、汎用性を高めるという点である。研究においては、独自性が重視されるが、一方で、研究が終了した後にデータを共有し、別の人がそれを、別の研究のための史料／データとして再利用することができるような基盤を形成していくことは、今後の歴史学の発展にとって重要である。

しかしながら、テキスト分析の手法をアラビア文字史料に応用する際には、もう一段階作業が必要となる。なぜならば、XML は左から右に書く文字を想定したものであり、右から左に書く文字に対しては極めて扱いにくいものであるためである。そこで、アラビア文字史料に対しては、より適したタグ付け手法を考案する必要があると同時に、先に述べた汎用性という観点から、TEI との互換性を担保する必要がある。講師は、現在この問題を検討するための共同研究を展開しているが、上記の問題が解決されれば、アラビア文字史料を用いたテキスト分析の研究が、汎用性や共有を理念に掲げる歴史情報学の中に適切に位置づけることができるという展望を持っている。