

漢文資料におけるアナログとデジタルの融合 ～その作成と利用と～

國學院大學兼任講師 篠原 泰彦

1. はじめに

1970年代に「パーソナルコンピュータ」と呼ばれる小型で低価格なコンピュータが発売されてから、約半世紀が経つ。1990年代に入ると、インターネットの普及により「情報端末」としての需要が高まり、携帯電話の高機能化や、スマートフォン、タブレットの普及がそれに拍車をかけた。平成31年4月から施行された学習者用デジタル教科書を制度化する「学校教育法等の一部を改正する法律」等関係法令や、令和元年に開始された文部科学省の「GIGAスクール構想ⁱ」による全生徒児童向けの学習者用コンピュータの整備もあり、教育の世界もデジタル化と無縁ではいられない。

しかしデジタル化は需要のある部分では急速に進む一方、企業活動や人々の日常生活から遠い部分では、その進展は個人の努力に依る部分が多い。漢文訓読のデジタル化については、インターネット上に個人のノウハウが散見されるものの、誰もが気軽に同じようにデジタル資料を作成できる、デジタル資料の作成法を新たに身につけるためのドキュメントが整備されている、という状況ではない。

また、漢文のデジタル化に関する論文の多くは、「返り点や送り仮名をどのようにデータとして扱うか」というマークアップ（構造化）への言及であり、教育現場ですぐに活用できる、あるいは資料の作成を容易にするものではない。ⁱⁱ

本発表では、最新の技術や手法を扱うのではなく、今ある技術やすぐ実践できる手法を、発表者がどのように用いてデジタル化を行っているかについて述べる。

2. 漢文訓読とデジタル資料

2-1. デジタル資料とは

本発表における「デジタル資料」とは、パソコン、タブレット、スマートフォンといったコンピュータで扱うことの出来る電子データを指す。JPEGに代表される画像（写真）ファイルや、Wordに代表される文書ファイル、Excelに代表される表計算ファイル、音声ファイルや動画ファイルなどが電子データの主なものである。

漢文の資料として用いられることが多いのは、Wordに代表される文書ファイルであろう。縦書き機能を持ち、漢文（原文）を入力・表示することができる。文字の大きさを変えたりルビを振ったりできることから、返り点や送り仮名の表記を行えるのも、周知の通りである。

2-2. PDF ファイル

忘れてはならないのが、PDF (Portable Document Format) ファイルである。PDF はコンピュータ上でアプリケーションや OS、ハードウェアに依存せず文章や図版を表示するために Adobe 社によって開発され、ISO 32000 で国際標準化されている。つまり PDF であれば、誰に渡しても文字化けせず、レイアウトが崩れることもなく、作成者の意図通りに表示することができる。漢文では正字を扱うことが多いため、文字化けの心配がないのは大きな利点である。

2-3. PDF ファイルの閲覧と作成

なお、以前は PDF を閲覧するために専用ソフトが必要だったが、現在はほとんどのコンピュータが標準機能で PDF を表示できるため、相手が PDF を閲覧できない可能性を考慮する必要はない。また、以前は PDF の作成には専用ソフトが必要だったが、現在は Office から直接 PDF を書き出すことが可能である。さらに、PDF にはしおりを付ける機能もあるため、適切に付けておけばページ数が多い場合でも目的の場所をすぐ閲覧することが可能である。

3. 漢文訓読に関するデジタル資料の作成

3-1. 漢文訓読に関するデジタル資料作成の障害

原文に返り点や送り仮名を付したデジタル資料を作成する際に問題となるのは、音読みや訓読みの変換では入力が難しい文字の入力と、返り点送り仮名の入力である。前者は日本語 IME (入力ソフト) の漢字検索機能や、異体字検索サイトの利用、フォントの自作などで、1文字ずつ入力することができる。後者はルビや下付き機能が利用されることが多い。どの方法が良いかについては本発表の主旨から外れるので言及しないが、どれも「紙に手で書くのに比べると手間がかかる」という点では共通している。しかし文字数が少なければ手書きの方が手間はかからないとはいえ、全て手書きすると当然時間がかかる。さらに、手書き資料は電子的に配信できない他、部分的な再利用が容易ではない。

3-2. アナログとデジタルを融合させたデジタル資料作成手法

そこで私は、デジタルを利用する方が楽な部分はデジタルを使い、デジタルを使うと手間がかかる部分は手書きで対応する方法をとっている。

まず原文は、可能な限りデジタルで作成する。これは、全てを手書きしようとする時間がかかることに加えて、インターネット上にデジタルデータが存在する場合は手元の原文との突き合わせが必要ではあるが、流用が可能だからである。その際、入力が難しい文字については無理に入力せず空けておく。

次に、作成したデジタル資料を紙に印刷し、空けておいた文字や返り点送り仮名を手書きする。手書きすることで、デジタルで作成するのに比べて所要時間を大きく減らすことができる。

最後に、紙をスキャナで読み取ってデジタル資料にすれば、電子的に資料を配付することができる上に、再利用も容易である。

3-3. スキャナを使用したアナログ資料のデジタル化

一般的なフラットヘッドスキャナを使う他、専用のドキュメントスキャナがあれば複数枚の原稿も短時間でスキャンできる。また、スマートフォンのアプリを使えば斜めから撮影しても正面から撮影したように加工してくれるため、綺麗にスキャンする事が可能である。

3-4. 紙を経由しない手書き資料の作成

手書きのできるパソコンやタブレット、スマートフォンなどがあれば、文字や返り点送り仮名を手書きする際に紙に印刷する必要がない。Adobe Acrobat や Office には手書き機能があり、スタイラス(電子ペン)を使えば比較的きれいに手書きすることができる。Office の「描画」タブには様々な形のペン先が用意されており、色や太さも調整できるため、紙に書く以上に豊かな表現をすることができる。

4. 漢文訓読に関するデジタル資料の利用

作成したデジタル資料を効率よく使用方法について、いくつか紹介したい。

4-1. サブモニタの利用

まず、サブモニタの利用である。デジタル資料を確認しながら作業を行いたい場合、ウィンドウを切り替えながらでは常に資料を見続けることができない。複数の資料を見比べたい場合も同様である。ウィンドウを並べる方法では、表示領域が小さくなり、全体を確認できないか、あるいは表示が小さくなってしまう。そのため、デジタル資料を紙に印刷している人も多いのではないだろうか。これを解消する方法の1つが、サブモニタの利用である。モニタをもう1台接続し、そちらに資料を表示する方法である。昨今はモニタの価格も下がっており、コンピュータをもう1台購入するよりかなり安い。

また、サブモニタを縦に使用方法もある。横長の画面で横書きであれば文字を大きく表示することができるが、横長の画面で縦書きの場合はそれができない。そこでモニタを縦置きにして、縦長の画面で縦書きを表示することで実現できる。縦置きに対応したモニタを購入せずとも、モニタアームを利用すれば多くのモニタを縦置きにできる。ただ縦に置くだけでは画面が横を向いてしまうが、OSで設定することで縦長に表示できる。

大きく表示することができればモニタを覗き込むことが減り、自然な姿勢で作業を続けられることから、疲労の軽減に繋がる。

4-2. 電子ペーパー端末の利用

次に、電子ペーパー端末の利用である。

電子ペーパーとは、微少な粒を任意の場所に移動させることで文字や画像を表示させる方法で、液晶のようにそれ自体が発光していないため、目に優しい。微少な粒を移動させるの

に時間がかかるため、表示の切り替えにはやや時間がかかり、動画の表示は不得手である。また、液晶の方が鮮明に表示できるほか、技術的にカラーの表示はまだ得意ではなく、写真のような鮮明さが求められる表示には適していない。しかし、文字の表示には十分であり、漢文のデジタル資料の表示には何の問題もない。また、液晶は発光させ続ける必要があるため電力を消費し続けるが、電子ペーパーは粒子の場所を移動させる時に電力を消費するため表示し続けてもほとんど電力を消費しない。そのため、スマートフォンやタブレットのような液晶端末に比べて、1回の充電で長く使用することができる。ただし、液晶のように素早く画面を切り替えることはできず、ページを切り替える時に一瞬待たされるため、素早くページをめくりながら該当箇所を探すような用途には適していない。

電子ペーパー端末は同サイズの液晶タブレットに比べると価格が高いものの、前述の通り発光しないため、目の疲れ方は紙と変わらない。そのため資料を長時間見続けることができる他、同サイズの液晶タブレットよりも重量が軽いため、持ち運びや手に持った長時間閲覧が容易である。

さらに、ガラス面への書き込みになる液晶端末に比べて、電子ペーパー端末への書き心地は紙へのそれに近く、デジタルにありがちな下手な文字になりにくいという利点もある。

4-3. プロジェクターの利用

最後に、プロジェクターの利用である。

モニタとして、プロジェクターを利用する方法もある。プロジェクターはスクリーンへ投影された光を目にするため、直接発光している液晶に比べて目が疲れにくい。また、設置場所次第ではあるが、40インチを超える大画面での表示を安価に実現することが可能である。昨今はホームシアター用に10万円を切る安価なプロジェクターが多数市販されており、ものによっては1万円程度でも購入できる。

ただし、プロジェクターの選定には注意が必要である。映画やゲームのように、細かい文字が表示されない用途であれば問題ないが、安価なものでは小さな文字が滲んで視認できないため、パソコンのメニューのような小さな文字でも視認できる高性能なものは最低でも20万円程度する。そのため、安価なものを使用するのであれば、大きく表示した文字を読み書きするため、と割り切る必要がある。逆に言えば、用途を選べば十分実用的である。

また、プロジェクターはモニタに比べて騒音が大きい他、自分の背後から投影する関係上、Zoomなどのオンラインミーティングでは前方にあるカメラの方向へ投影することになるため、使用が難しいことを申し添えておく。

5. 漢文訓読に関するデジタル化の未来

ここまで紹介してきた様々な技術は、日々進化している。

5-1. TEIによる漢文訓読ガイドライン策定

例えば、人文学資料をデジタル化するためのガイドライン策定を目的とした、TEI（テキスト・エンコーディング・イニシアチブ、Text Encoding Initiative）という国際的な共同

プロジェクトがあり、TEI コンソーシアムによって運営されている。ⁱⁱⁱ 日本ではこれまで個人的な関与に留まっていた^{iv}が、2016 年には日本語資料に対応するための環境を整備する分科会である東アジア／日本語分科会 East Asian/Japanese Special Interest Group が作られ、漢文訓読への応用も始まっている。漢文訓読のガイドラインが策定されれば、それに対応したソフトウェアの開発に繋がり、デジタル化はさらに容易になるであろう。

TEI と人文学との関わりについては『人文学のためのテキストデータ構築入門: TEI ガイドラインに準拠した取り組みにおいて』^{vi}が詳しい。

5-2. OCR によるアナログ資料のテキストデータ化

また、既に作成されている紙資料からデジタル資料を作る方法として、OCR（光学文字認識、Optical character recognition）という方法もある。これはアナログ資料を画像としてデジタルデータにするのではなく、アナログ資料の文字をテキストデータとして認識させる技術である。手書き文字の認識は不得手であり、活字であっても何の調整もせずに使えるほどの精度を持つものはないが、元になる資料によるものの調整や確認の時間を勘案しても入力の手間が半分程度になることも多い。

市販ソフトやスマホアプリも存在するが、手軽かつ比較的認識率が高いのは、Google ドライブである。Google ドライブに JPEG ファイルもしくは PDF ファイルをアップロードして、それを Google ドキュメントで開くと自動で文字を認識する。ただし、Google ドライブへのアップロードは、著作権と情報漏洩に注意が必要である。

6. おわりに

これまで述べてきたように、デジタル化が難しいという印象の強い漢文訓読資料だが、実際にはデジタルだけにこだわらずアナログと上手く組み合わせることで、手間をかけずにデジタル化することが可能である。

本発表を振り返ると、まず PDF ファイルによって誰でもどのような環境でも文字もレイアウトも同じように表示できることに触れた後、以下の 3 点について言及した。

1. 漢文訓読に関するデジタル資料作成には、入力の難しい文字の存在や、返り点送り仮名の入力が煩雑という問題がある。
2. デジタル化が難しい部分は割り切って手書きで行うことで時間と労力を削減できる。
3. 紙に印刷したものに手書きしてスキャナでデジタル化するか、デジタルデータへ手書きできる電子機器を活用してデジタルのまま処理する方法がある。

次に、デジタル資料の閲覧は、ただパソコンなどで閲覧するだけではなく、より適した閲覧方法を採用することで、作業効率を高めることができる。これについては以下の 3 点について言及した。

1. サブモニタを活用する。またそのサブモニタを縦置きで利用する。
2. 電子ペーパー端末を利用する。

3. プロジェクターを利用する

最後に、「漢文訓読に関するデジタル化の未来」として TEI と OCR について述べた。

技術の進歩によって負担が増えるのは本末転倒であり、本来は私たちを助けてくれるために進歩するものである。技術を上手く使い、労力を軽減し、あるいは同じ労力でより良いものを作成するために、本発表がその一助となれば幸いである。

ⁱ GIGA スクール構想について：文部科学省

https://www.mext.go.jp/a_menu/other/index_0001111.htm

ⁱⁱ 「コンピューター上の漢文訓読表記法に関する現状と課題」篠原泰彦、國學院中國學會報第 65 輯、2019 年、國學院大學中國學會

ⁱⁱⁱ TEI: Text Encoding Initiative <https://tei-c.org/>

^{iv} TEI Day in Kyoto 2006 <http://coe21.zinbun.kyoto-u.ac.jp/tei-day/abstracts.html.ja>

^v 「デジタルアーカイブ」におけるテキスト資料の構造化・マークアップの国際デファクト標準、TEI ガイドラインを作っている TEI 協会に日本語資料を扱う分科会ができました - digitalnagasaki のブログ <https://digitalnagasaki.hatenablog.com/entry/2016/06/30/030347>

^{vi} 『人文学のためのテキストデータ構築入門 TEI ガイドラインに準拠した取り組みにおけて』一般財団法人人文情報学研究所監修、石田友梨／大向一輝／小風綾乃／永崎研宣／宮川創／渡邊要一郎（編）、2022 年、文学通信

《 資料 》

○ 異体字検索サイトの利用

文字を入力して「検索」することで、入力した文字の異体字を検索することができるサイト。検索した異体字はコピー＆ペーストで使用することができる。

ただし、ものによっては特定のフォントでなければ表示されない場合もある。コピー＆ペーストで上手く表示されない場合は、フォントを変えてみると良い。

- 異体字リスト | 東文研 総合検索

<https://www.tobunken.go.jp/archives/%E7%95%B0%E4%BD%93%E5%AD%97%E3%83%AA%E3%82%B9%E3%83%88/?moji=%E5%AD%A6#>

- 文字情報基盤検索システム

<https://moji.or.jp/mojikibansearch/basic>

- GlyphWiki:異体字 - GlyphWiki

<https://glyphwiki.org/wiki/GlyphWiki:%E7%95%B0%E4%BD%93%E5%AD%97>

- 《Unicode/CJK 統合漢字》漢字検索

<https://rtk.sakura.ne.jp/cjk/>

あらゆる漢字が部首別もしくは総画数別に並んでいる中から自分で探す。

○ フォントの自作

- GlyphWiki:異体字 - GlyphWiki

<https://glyphwiki.org/wiki/GlyphWiki:%E7%95%B0%E4%BD%93%E5%AD%97>

- 「I 文字フォント」を作成することができる
 - ☆ 別の漢字の一部を参照しながら文字を作成できる
 - ☆ 太字は作れない
 - ☆ 書体は選べない
 - ☆ 通常の文書を他者に渡す場合は、I 文字フォントも渡してインストールしてもらう必要がある（PDFであればフォントが埋め込まれるのでそのまま渡しても問題ない）

○ 再読文字の実現はハードルが高い

- 再読文字の入力：Word で漢文入力

<https://www.shuiren.org/chuden/toyoshi/kanbun/03.htm>

- Word の場合「フィールドコード」を使用することで実現できるが、一種のプログラミングのため難易度が高い。また、Word に負荷がかかるため、1つの文書で複数回用いると Word がほとんど操作できなくなってしまう場合もある。
- 再読部分（原文の漢字の左側）については、テキストボックスで実現する方法が手軽だが、行や文字の挿入や削除で原文の漢字の場所が移動してもテキストボックスは付随して移動しないため、再読部分が正しい場所にあるか常に確認が必要。

○ 原文サイトの利用

原文をI文字ずつ入力するのは労力がかかるため、以下のようなデータベースを利用して、原文をコピー&ペーストして使用することも効率化には有効である。

ただし、原文データの正しさは担保されていないため、使用する際は必ず確認が必要である。

- 中国哲学書電子化計画（<https://ctext.org/zh>）
「哲学」と題するが、文・史を含めた先秦から民国までのあらゆる電子テキストを公開している。
- 中央研究院「漢籍電子文献」（<http://hanchi.ihp.sinica.edu.tw/ihp/hanji.htm>）
台湾中央研究院が提供する漢籍電子文献。
(免費使用)と(授權使用)とに分かれるが、個人での申請資格は台湾在住の人に限られる。前者だけでも、『史記』から『明史』までの二十四史や『清史稿』など、基礎的な古典文献を検索・閲覧できる。
- 維基文庫（<https://zh.wikisource.org/>）
ウィキペディアの姉妹サイト、ウィキソースの中国語版。
比較的多くのテキストデータが収集されているが、底本を明記していないデータも多い。

- 漢籍リポジトリ (<https://www.kanripo.org/>)
『四部叢刊』・『四庫全書』・『正統道蔵』などのテキストデータが収録されている。
日本語のサービスなので見やすい。
- 寒泉 (<http://skqs.lib.ntnu.edu.tw/dragon/>)
十三経・二十四史・『清史稿』などのほか、『全唐詩』・『資治通鑑』などが検索できる。
異体字の一括検索に対応していないなどの問題もある。
- 宋詩 (<http://cls.lib.ntu.edu.tw/QSS/home.htm>)
台湾・元智大學の羅鳳珠氏のページ。
蘇軾・陸游・楊萬里・梅堯臣・歐陽修・王安石などの詩の検索ができる。

○ 主な電子ペーパー端末

• Amazon

Amazon の電子書籍閲覧端末としても利用できる

- **Kindle Paperwhite** (<https://www.amazon.co.jp/dp/B08N41Y4Q2>)
 - ◇ 画面サイズ 6.8 インチ / 約 1.7 万円
 - ◇ PDF ファイルの閲覧はできるが、書き込みはできない
- **Kindle Scribe** (<https://www.amazon.co.jp/gp/product/B09BRG33C3>)
 - ◇ 画面サイズ 10.2 インチ / 約 5 万円
 - ◇ PDF ファイルへ手書きによる書き込みが可能

• 富士通

- **クアデルノ** (<https://www.fmworld.net/digital-paper/top.html>)
 - ◇ A4 サイズ (約 7 万円) と A5 サイズ (約 5 万円) がある
 - ◇ A4 サイズであれば、多くの資料を原寸大で表示できる

• Onyx

- **Boox** (日本での販売代理店：SKT 株式会社 <https://sktgroup.co.jp/>)
 - ◇ 液晶ではなく電子ペーパーを搭載した Android タブレット
 - ◇ Android タブレットなので、アプリストア (Google Play) から様々なアプリをインストールして使用可能

BOOX Tab X	13.3 インチ	約 13 万円	A4 サイズ
Tab Ultra	10.3 インチ	約 8.5 万円	
Tab Ultra C	10.3 インチ	約 9 万円	カラー
Nova Air2	7.8 インチ	約 5 万円	
Nova Air C	7.8 インチ	約 9 万円	カラー
Leaf2	7 インチ	約 3.5 万円	
Poke5	6 インチ	約 2.5 万円	