

漢籍デジタルテキストの可能性

—『文苑英華』の校勘を通じて—

はじめに 書状関連史料を蒐集する作業の一環として現在、日本の史料のみならず漢籍史料に対象を広げて作業を遂行している。その際近年急速に普及した漢籍のデジタルテキストの利用を試み、さらにデータの校勘作業をおこなっている。実際の作業の中で新たな興味深い事実が発見できたので紹介したい。以下の執筆は作業にあたった山元がおこなう。

漢籍テキストデータとフォント 我々は現在、宋代に編纂された詩文集『文苑英華』に収録されている唐代の書状(表・牋・狀・啓)について、『文淵閣四庫全書電子版』所収のテキストデータを、中華書局影印本『文苑英華』と比較する校勘の作業をおこなっている。

校勘作業はまず、『電子版』のテキストデータをコピーし、一首ずつに切り分けて、Excelの表に貼り付けるところから始まった。しかし作業が始まった数年前、Excelに貼り付けられた本文データは、表示できない文字を示す「？」が多く含まれるものだった。当時使用していたPCのOSはWindowsXPであったが、Excelで使用するフォントが日本語入力用のフォントであるMSPゴシックに設定されていたため、日本語ではあまり使われない漢字を表示できなかったものと思われる。

現在使用しているPCのOSはWindows10であるが、日本語用のフォントとして表示できる漢字の数が増え、表示できない字体もSimSunなどの中国語用フォントに置き換えられるため、Excelで同じ作業をしても、表示できない漢字はかなり少なくなった。WindowsXPにもSimSun(簡体字)MingLiU(繁体字)などの中国語用フォントは標準搭載されていたので、MingLiUをExcelで使用するフォントとして設定した上で、本文のテキストを貼り付けていれば、「？」の数は最小限に抑えられたものと思われる。「？」が本来どの文字だったかという情報は失われていたので、本文データの貼り付け作業は、一部を除きやり直すことになってしまった。

我々と同じ文系の研究者の中にも、PCによるデータ作成や、データベースの構築などに精通した人はいるが、我々と同様、何が問題の本質なのか見当もつかないまま、時間を空費している人も多いであろう。まさに「少

しの事にも、先達はあらまほしきことなり」である。

一直線の異同の謎 本文テキストのExcelへの貼り付けが終わると、いよいよ中華書局本『文苑英華』との対校の作業に入る。中華書局本『文苑英華』の我々が検討の対象とした部分のうち、巻533～600は明の隆慶元年(1567)刊本の影印、巻601以降は宋本、つまり『文苑英華』の編纂後、最初に印刷された刊本の影印である。現在対校の作業は明本の部分までで、巻601以降の宋本部分には至っていない。四庫全書本と中華書局本の本文を一首ずつ、上下一組にしてExcel上に配置し、異なる部分を強調表示するという手法を取っているが、多くは一文字単位の異同である。単純な誤りによるものもあれば、諱字(皇帝やその父祖の名に関わる字)を避けるため、あえて書き換えられたものもある。

四庫全書本と中華書局本の文が大きく食い違う部分も、少数ではあるが見つかった。図1に挙げた巻569「為杭州崔使君賀加尊号表」では、中華書局本(明隆慶元年刊本)の各行の中ほどに、四庫全書本とはまったく異なる文字が3文字ずつ、横一直線に並んでいる(網フセ部分)。

巻569部分の宋本は現存しないが、景宋鈔本と呼ばれる、宋本に紙を乗せトレースしたものが残っており、隆慶元年刊本の「為杭州崔使君賀加尊号表」の文章は、宋本のものとほぼ同じであることがわかる¹⁾。それにしても、なぜ隆慶元年刊本と四庫全書本との間に、このようなかたちの異同が生じたのだろうか。

問題の3文字の上と下の部分(図46の◀)をよく見ると、界線が途切れ、上下の文字の間隔がやや開いて、白い横線が入っているように見える。版本のこの部分に、木目に沿った横の割れ目があったのであろう。

明本の『文苑英華』は隆慶元年の初版の後、隆慶6年・万暦6年(1578)・万暦36年に、傷んだ版本の補修と再版がおこなわれている(遞修本)。四庫全書本の底本となったのは、このうち万暦年間に再版された版本であったとされる²⁾。初版から再版までの間に、図1部分の版本も破損が進み、割れ目に挟まれた3文字の部分が失われたのであろう。そこに何者かが、前後の文字から類推して、本来のものではない文字を埋めたわけである。

「国立国会図書館デジタルコレクション」³⁾所収の『文苑英華』1000巻は、隆慶6年・万暦6年・万暦36年の補修・再版を経た遞修本の画像データである。その「為杭

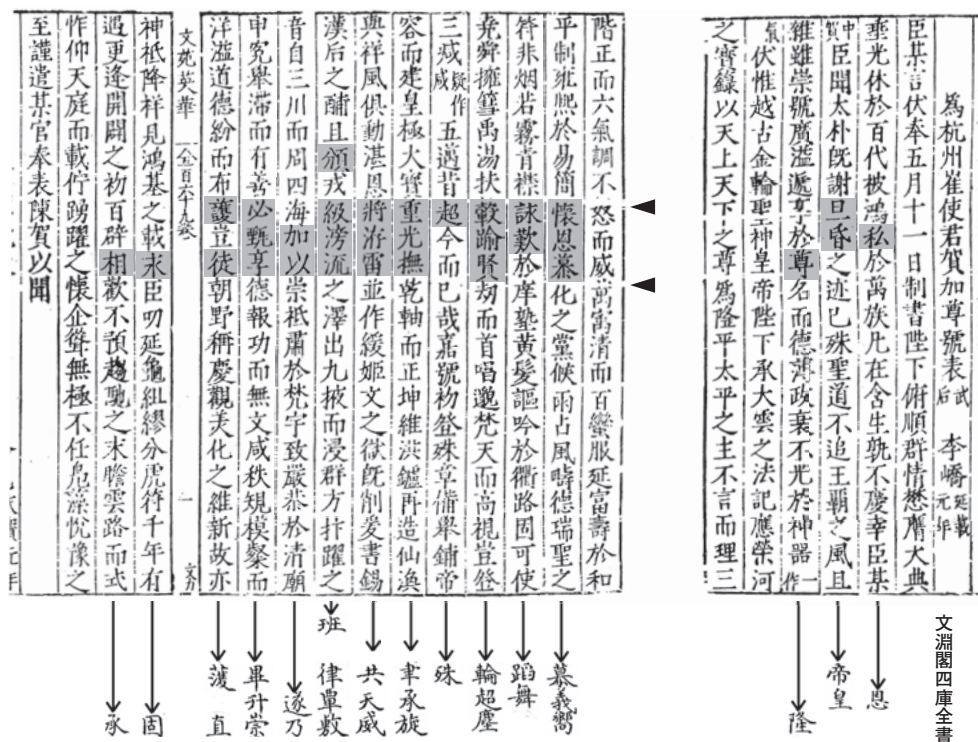


図46 『文苑英華』 卷569「為杭州崔使君賀加尊号表」

州崔使君賀加尊号表」の部分を見ると、図46部分に対応する2葉の中央の折り目には、隆慶元年刊本にはない「三十六年刊」の文字がある。この2葉の版木（おそらく同じ板の表と裏）は損傷が激しかったため、万暦36年の再版時に新たに彫り直されたのである。問題の3文字の部分には、隆慶元年刊本と同じ文字（ただし「於」の字体が異なる）が入っている。何らかの底本を元に、本来入るべき文字を調べ、補ったのであろう。

そうすると、四庫全庫本の「為杭州崔使君賀加尊号表」部分の底本は、万暦36年より前、隆慶6年か万暦6年の遞修本であった可能性が高くなる。では3文字の部分に適切な文字を埋めたのは誰か。明の遞修本の制作に関わった者か、それとも四庫全書の制作に関わった者か。

四庫全書は清の乾隆年間に、印刷ではなく手書きで、計8部作成された。そのうち文淵閣に収められたものが、今扱っている『電子版』の底本であるが、それとは別の、文津閣に収められた四庫全書の影印本⁴⁾を確認してみると、『文苑英華』「為杭州崔使君賀加尊号表」の問題の3文字部分には、文淵閣本とも隆慶元年刊本とも異なる、独自の文字が入っているのである。このような3資料間の異同は、巻572「為王及善讓内史第二表」「为王方慶讓鳳閣侍郎表」にも見られる。問題の3文字部分を推測で埋めたのは、どうやら四庫全書の筆写に携わる者だった可能性が高そうである⁵⁾。

まとめ 『文苑英華』を含む『四庫全書』の全文テキストデータには、web上で無償で公開されているものもある。「中国哲学書電子化計画」⁶⁾「漢籍リポジトリ」⁷⁾の両

サイトのものには、底本の画像データも添えられている。

一つの文献について諸本を比較検討し、本来の姿に近づける校勘の作業は、漢籍研究の非常に長い歴史の中で、繰り返し営々とおこなわれてきたものである。ここで述べた「為杭州崔使君賀加尊号表」の文の異同の件なども、とうの昔に知られていたことであろう。しかし近年の状況がこれまでと異なるのは、書庫に入り浸ることのできる一部の研究者だけではなく、様々な地域の、様々な立場の人々が、webを通じてテキストデータを検索し、画像を見るなどして、膨大な漢籍に効率的にアプローチできるようになったということである。底本やデータの質、省かれてしまう情報の多さなど、改善すべき課題は多いだろうが、今後の技術の進展と、より多くの人々に開かれた研究環境の実現に期待したい。

なお、本稿は、JSPS科研費JP24520749「書状文化の源流を求めて」（平成24～27年度、代表者黒田洋子）の成果の一部である。

（黒田洋子／客員研究員・山元章代／奈良女子大学）

註

- 1) 傅增湘『文苑英華校記』北京図書館出版社、2006。
- 2) 凌朝棟『文苑英華研究』上海古籍出版社、2005。
- 3) <http://dl.ndl.go.jp/>
- 4) 『文津閣四庫全書』商務印書館、2005。
- 5) 中国国家図書館には『文苑英華』の隆慶6年・万暦6年遞修本とされる本が所蔵されている。文淵閣・文津閣四庫全書の底本はこれと同じ版であった可能性があるが、今回その画像や複写を入手することはできなかった。
- 6) <http://ctext.org/zh>
- 7) <http://www.kanripo.org/>