

漢字文献のオンラインデータベースについての一考察

山崎直子

Naoko YAMASAKI : A Study of Online Chinese Text Database

中国語の古文献を用いて目的の表記を見つけだすには、かつては厚い書物を1ページずつめくっていくのが当然であった。近年、コンピュータの普及に伴い、古文献の電子テキスト化が進んできた。また、インターネットの普及により、電子テキストをオンラインで利用することも可能となっている。特に若い世代では、インターネットでの検索というのはポピュラーな手段であり、研究に利用できる。データベースの作成には諸問題もあるが、利用可能なオンラインデータベースも構築されている。

キーワード：漢字文献、オンラインデータベース、電子テキスト

1. はじめに

筆者の大学の卒業論文は「中国古典に見える『鬼』について」であった。春秋左伝、国語、礼記、周礼、儀礼の中から「鬼」についての記述を探し出し、その意味するところを探った。

春秋左伝は魯の隠公元年 (B.C.722) から哀公十四年 (B.C.481) までの242年とその2年後、孔子の死去するまでを年代を追って書かれた歴史書であるが、書物として相当のページ数がある。それを隅から隅まで読み、「鬼」についての記述を拾い上げ

ていく。さらに、それぞれの記述について大漢和辞典をめくりつつ解釈を行った。研究者であれば当然の過程であるが、学生であった当時は大変な作業であったと記憶している。

近年では、インターネットの普及により、家庭でも学校でも情報検索・収集が容易になっている。平成18年度の総務省による調査 (図1) では、10~20代のインターネット利用率は93~94%と、ほかの世代に比べても大変高い。実際、レポートの課題が出れば、多くの学生はインターネットを立ち上げ、検索バーに課題タイトルを入力するだろう。

かつて私がそうであったように、中国語に堪能で

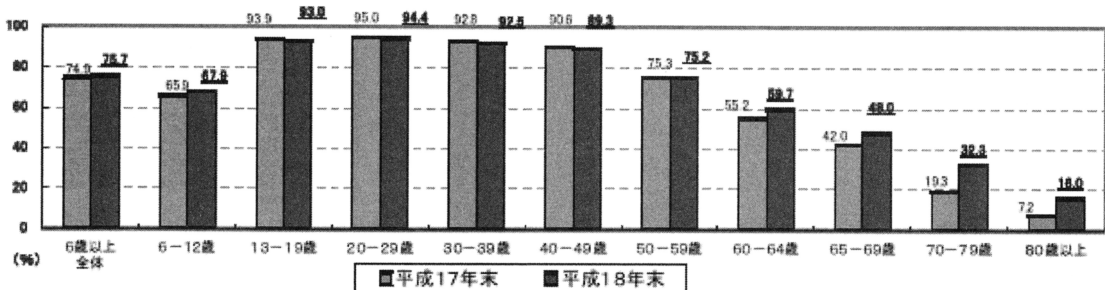


図1 世代別インターネット利用率¹⁾

ない学生が古文書を利用し研究を行おうとしたとき、インターネット経由で全文検索できるデータベースが利用できれば、ずいぶん便利ではないか。

多くの学生は、インターネットで検索機能などが使えるが、コンピュータに精通しているわけではない。また、皆が中国語が堪能というわけでもなく、中国語で記述されたページをすべて理解することはできない。本稿では、オンラインで漢文の全文検索システムをこういった学生の立場で利用することを想定して、その利便性について考えていきたい。

2. オンラインデータベースの検証

対象サイトは中國哲學書電子化計劃²⁾と漢籍電子文獻³⁾である。

前者はイギリス人のDonald Sturgeon氏作成によ

るもので、自身の研究の際、中国の古文書の電子テキストの少なさを痛感し、その解消を目的に作成されたものである。Wikipediaの「春秋左氏伝」のページからリンクがあるため、学生がたどり着きやすいと考えられる。簡体字中国語（文字コード：GBコー

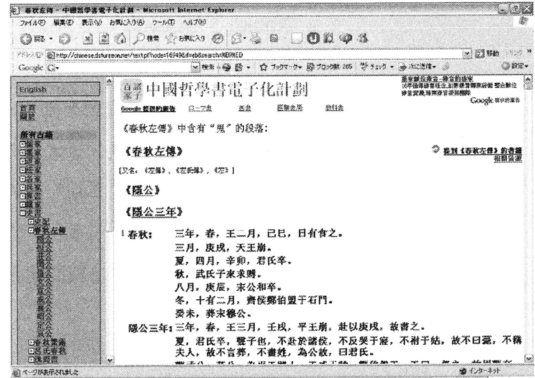


図4 中國哲學書電子化計劃検索結果3



図2 中國哲學書電子化計劃検索結果1

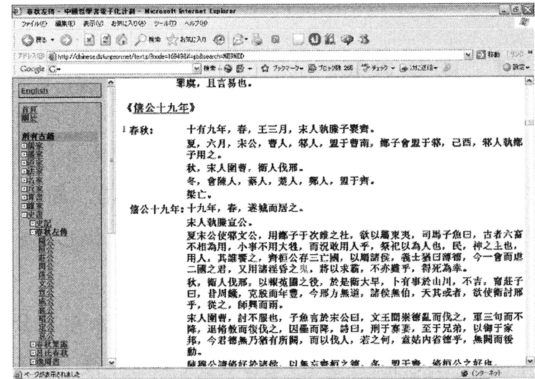


図5 中國哲學書電子化計劃検索結果4

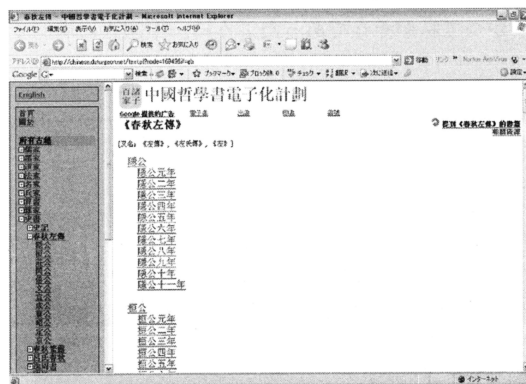


図3 中國哲學書電子化計劃検索結果2

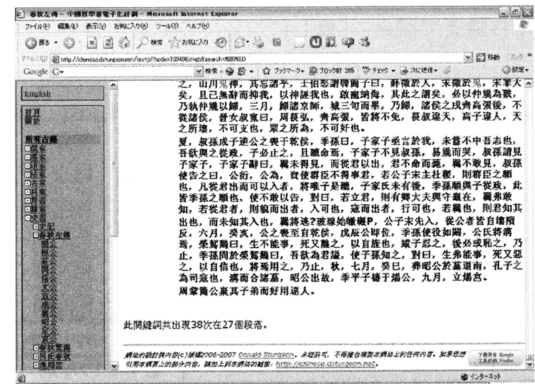


図6 中國哲學書電子化計劃検索結果5

ド)を用いており、英語版もある。

後者は台湾の国立研究機関である中央研究院のWebサイト上のデータベースである。研究者にはすでにポピュラーなサイトである。繁体字中国語(文字コード:Big5コード)を用いている。

(1) 中國哲學書電子化計劃

まずは中國哲學書電子化計劃で実際に検索を行った。

所有古籍メニューから春秋左伝を選択, 入力ボックスに「鬼」と入力し, 検索をかける。(図2, 3) 検索結果は段落で表示され, 目的の文字は赤で表示される。(図4, 5) 最下部には検索結果として表記の数と段落数が表示される。(図6)

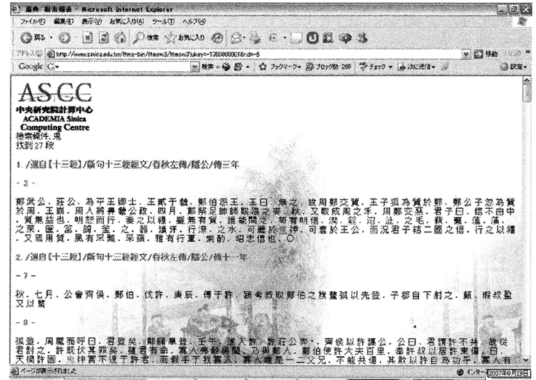


図9 漢籍電子文獻検索結果3

(2) 漢籍電子文獻

次に漢籍電子文獻についてである。

トップから十三経→十三経注・(一八一五年阮元刻本)→春秋左傳正義と移動する。その後, 検索条件として「鬼」と入力し, 春秋左傳正義にチェックを入れ執行ボタンをクリックする。(図7)

検索結果はまずは段落単位で表示され, 上部には検索結果として段落数が表示される。目的の文字は赤に下線で表示される。(図8) また, 全段落表示を選択することも可能である。(図9)

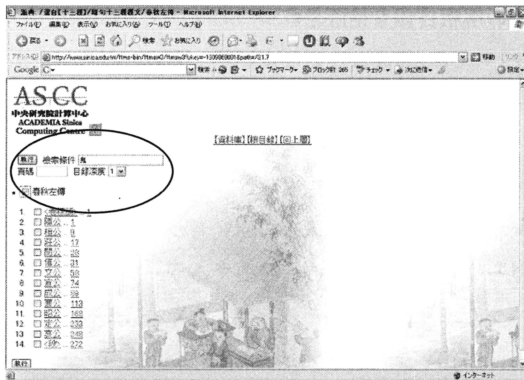


図7 漢籍電子文獻検索結果1

(3) 異体字の検索

次に多くの古文獻では異体字がつかわれる「体」を検索してみた。「体」は多くの古文獻では異体字「體」で表現される。

中國哲學書電子化計劃では, 異体字「體」が検索結果として表示された。(図10) 一方, 漢籍電子文獻では, 検索エラーが表示された。

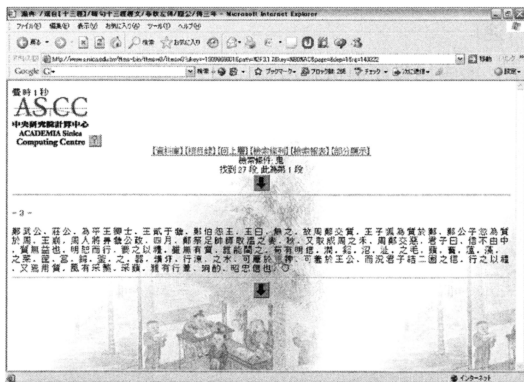


図8 漢籍電子文獻検索結果2

(4) 文字化け

今回の検索結果では対象段落中にいくつかの文字化けが表示された。中國哲學書電子化計劃では隠公三年に図10下線部のような表示があった。本来は, 「去順效逆。所以速禍也。君人者將禍是務去。而速之。無乃不可乎。」⁴⁾と記述されている部分である。

また, 漢籍電子文獻にも文字化けがあった。

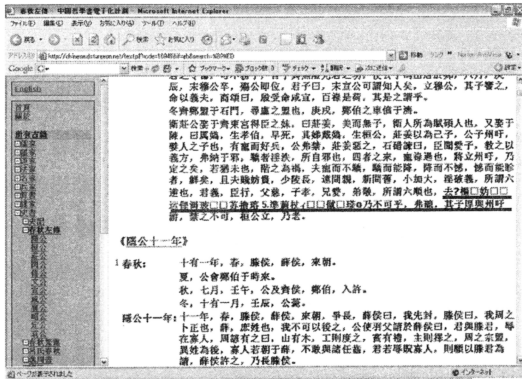


図10 中國哲學書電子化計劃（異体字）

3. 検証結果について

以上のように、「春秋左伝」内の記述について2つのオンラインデータベースを検証した。

中國哲學書電子化計劃は作成者が英国人ということもあり、中国語以外にも英語での表示を選択することが可能である。英語に関してはほとんどの学生が何らかの学習を経ているので、若干わかりやすい。また、カテゴリわけも対象文献が限られているためわかりやすいものとなっている。

一方で漢籍電子文献では、文献数も増加中で、大きなカテゴリわけから始まっている。メニューとして一覧が表示されるわけではないので、戸惑うこともあるのではないだろうか。また、一部は有料となっているのですべての利用が可能ではない。

しかしながら、いずれのサイトも入力ボックスに検索文字列を入力し、検索ボタンをクリックすれば検索結果が表示された。インターネットの検索機能を利用したことのある者なら、まず問題ないだろう。つまり、学生の立場で表記を検索するためには、十分有効な利用ができるとみられた。多少の文字化けなどはあったが、書物と併用すれば十分解消される問題である。もちろん、作成者も日々修正を加えており、本稿執筆中に、上で指摘した中國哲學書電子

化計劃の文字化けは解消されていた。

文字コードもBig5コードあるいはGBコードが使用されていたが、少なくともInternet Explorer 6では、自動コード選択でほぼ問題なく閲覧が可能だった。マーケットシェア⁵⁾調査によると、2007年9月のInternet Explorerのシェア率はおよそ78%で、依然ほかのブラウザをと大差で1位であり、利用者のニーズには応えられていると言える。

4. おわりに

漢字文献のデータベースは、日々作成が進んでいる。特に外国の研究者の手によるものが多く、日本人の学生、日本語以外に堪能でない者にとっては若干の不便が生じる。また、甲骨文、金文などはもともと紙ベースではないこともあり、テキストではなく画像としてのデータベースになる。また、その文字の解釈自体も見直されつつあり、文字検索用データベースの作成はなかなか困難であるのが現状である。

従来の方法で、ぶ厚い書物を1ページ1ページめくって行ってこそ研究だという考えも依然あるが、時代の流れに乗り、未来の研究者が研究に取りかかりやすい状態を作るのも必要だろう。

また、今後資料を残していくためにも紙ではなく、劣化しない電子テキストという形を作り上げることも重要である。

(注)

- 1) 平成18年度通信利用動向調査 総務省
- 2) http://chinese.dsturgeon.net/index_gb.html
- 3) <http://www.sinica.edu.tw/ftms-bin/ftmsw3/>
- 4) 漢文大系 第十卷 左氏會箋上 富山房、1990
- 5) <http://marketshare.hitslink.com/>