Vol.2018-CH-117 No.2
2018/5/12

加点情報の再構成

高田智和[†1]　田島孝治[†2]　堤智昭[†3]

概要：ヲコト点移点ツールを作成し，実際の訓点資料（国語研蔵『尚書（古活字版）』）を対象として，データ作成を行った．訓点（ヲコト点）の定量的研究を行うための基礎データを得る見通しが立った．

キーワード：訓点資料，漢文訓読，ヲコト点

Reconstruction of *Kunten* Materials for Quantitative Analysis

Tomokazu TAKADA[†1]　Koji Tajima[†2]　Tomoaki TSUTSUMI[†3]

Abstract:.For quantitative analysis of *Kunten* Materials, we developed the *Iten* tool and constructed a data of wokoto-ten information.

Keywords: *Kunten* Material, wokoto-ten

1. はじめに

かつて漢文は社会・文化・学術を語る東アジアの共通書記言語であった．東アジアの中国周辺民族は漢文を受容し文明を育んできた．日本も例外ではない。自言語とは異なる漢文を受容し理解するにあたって，東アジア各地で漢文訓読（自言語による理解）が行われた．最も高度に発達し長く続いたのが日本の漢文訓読である．

漢文訓読のために訓点を記入した典籍・文書を訓点資料と言う．訓点資料は，前近代の政治，経済，思想，宗教，文芸などあらゆる分野の典籍・文書に及び，歴史学，文学，仏教学，文化学など広く史的研究を行う上での根幹資料となっている．また，日本語史研究では，文字（片仮名の成立，仮名遣の確定），音韻（音便，漢字音），語彙・語法などの史的変遷を解明するために，訓点資料が活用されてきた．

2. 訓点資料研究における課題

日本語史研究の中での訓点資料研究には，現在いくつかの課題がある．

9 世紀平安時代初期においては信頼に足る和文系資料が存在しないため，訓点資料を用いた種々の研究成果が公表されてきた．大矢透[1]，春日政治[2]，中田祝夫[3]，築島裕[4]などである．訓点資料から，平安時代初期以降の日本語諸現象の事実と変遷とが見出され，語彙や文体の面で和文系資料とは異なる漢文訓読語彙・漢文訓読体が，歴史的に併存してきたことが明らかとなった．一方で，ヲコト点を中心とする訓点の研究は高度に発達し，あるヲコト点を使用する教学流派に対する理解なくして訓点資料研究は成り立たなくなっている。このことは，訓点資料の専門研究者でない日本語史分野の研究者が，訓点資料の研究利用を敬遠する一因になっていると考えられる．

また，日本語史研究では定量的分析が盛んである．しかし，訓点資料研究では，網羅的かつ大規模な訓点の定量的分析は存在しない．語種比率や品詞比率など，和文系資料において既に計測されている基礎計量データですら，訓点資料から容易には得られない．この背景には，本文が漢文であることに加え，訓点は複雑な書き入れであるため，訓点資料研究へのデジタル技術導入の立ち遅れがあると考えられる．和文系資料には「日本古典文学大系本文データベース」[5]や「日本語歴史コーパス」[6]があり，定量的分析を行うための装置が整っていることと比べて，未だ手作業に依られねばならない訓点資料との環境差は歴然としている．

3. 訓点の定量的分析を目的とした電子化

筆者らは，訓点の定量的分析を目的として，訓点資料の電子化について取り組んできた（高田智和[7]，田島孝治ほか[8]，堤智昭ほか[9]，林昌也ほか[10]）．韓国の口訣資料研究で採用されている座標表現（李丞宰[11]，朴鎭浩[12]など）を日本のヲコト点に援用し，ヲコト点図集収載のヲコト点を電子化して，基礎計量をするに至った（堤智昭ほか[13]）．ヲコト点情報を電子化するための環境は整ったものと考えられる．

ヲコト点図集収載のヲコト点は，いわば文字コード表の

†1 国立国語研究所
National Institute for Japanese Language and Linguistics
†2 岐阜工業高等専門学校
National Institute of Technology, Gifu College
†3 東京電機大学
Tokyo Denki University

ⓒ 2018 Information Processing Society of Japan

ようなものであり，文字コード表の文字を調べても，文字の使用実態を知ることはできない．実際の訓点資料に記入された訓点を調べなければ，訓点資料に反映された言語使用の実態を記述することはできない．ヲコト点図集収載のヲコト点電子化の知見をもとに，いよいよ実際の訓点資料の訓点記述を本格化することにした．

訓点資料の解読では，移点，釈文作成，書き下し文（いわゆる訓読文）作成の 3 段階を経る．第 1 段階の移点は，原本，あるいは，原本に準じる複製，影印，デジタル画像を見ながら，訓点の有無を確定していく作業である．具体的には，漢文本文を用意してそこに訓点を記入したり，マイクロフィルムの紙焼き等にマークを入れたりする．現存する訓点資料も，第 1 回目の加点資料ではなく，移点資料であることも少なくない．

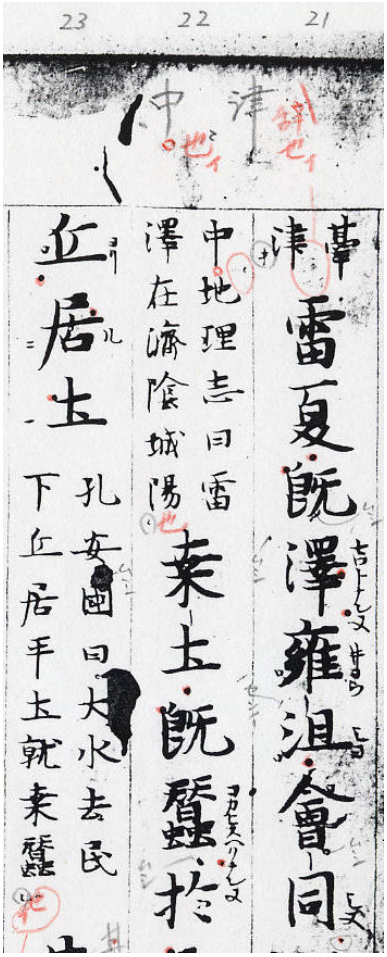

図 1　移点結果

第 2 段階として，移点結果をもとに釈文を制作する．釈文は，漢文本文の体裁を保ったまま翻字し，ヲコト点を仮名にするなど一定のルールを設けて訓点を書き込んだ翻刻文である．初期解読の成果と呼べるものである．

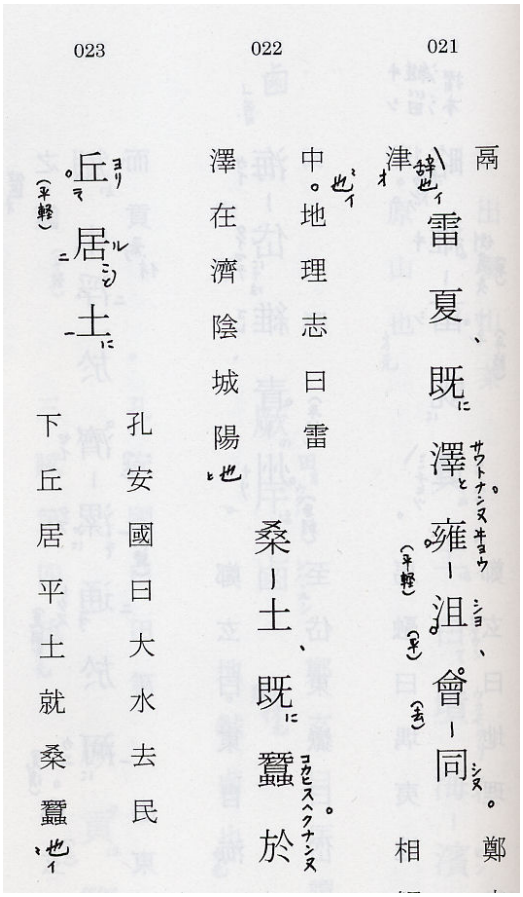

図 2　釈文

第 3 段階として，移点結果，あるいは釈文から書き下し文を制作する．書き下し文にも幾通りかあり，中でも，現在の学校教育で行われている，原漢文の語順を日本語文（古文）に直す方法が一般的である．日本語史研究の書き下し文では，原漢文の語順のまま書き下し文を作る方法もある．日本語としての文レベルでの文法・語法研究を意図するならば，原漢文の語順を日本語文に直した書き下し文が便利であろう．釈文であっても，文字あるいは語レベルでの研究利用は可能である．

訓点資料の訓点そのものを研究対象とする場合には，まずは第 1 段階の移点結果を電子化対象とするのが良いだろう．したがって，次に述べる訓点情報電子化のためのツールは，移点支援のためのツールと呼べるものである．

4.　訓点情報電子化のためのツール

4.1　対象とする訓点の種類

今回電子化の対象とした訓点の種類と意味を表 1 にまとめる。また，資料は国立国語研究所蔵『尚書（古活字版）』9 巻[14]である．『尚書（古活字版）』に存在するすべての訓点情報を電子化し計量することが理想ではあるが，資料に精通した者でなければ困難である．特に，仮名点や文字を含む校正符号などを入力することは，字形解読が重要とな

ⓒ 2018 Information Processing Society of Japan

ることから，困難である．この実現には，画像処理などの他の技術との組み合わせが必要であろう．そこで，今回は加点の原理を理解していれば容易に入力できるであろう，訓点記号を選んだ．これらは，記号の位置，記号の形状と，朱や墨などの色彩情報を入力できれば，電子化することが可能である．

表 1　電子化する加点情報の種類

訓点の名称	訓点の持つ意味・解釈
科段点	段落
句点	文
読点	文中の区切り
合符	語のまとまりや解釈（人名を表す符号も含む）
声点	語のアクセント（語の解釈の確定）
ヲコト点	助詞・助動詞など（語形の一部の場合も含む）

4.2　ツールの仕様

該当する訓点情報を入力するには，色彩や形状を選び，本文の文字の上で座標を選んでいくと効率が良い．また，どの位置に訓点情報を入力したかが一目で分かると入力ミスも少ない．今回制作した入力ツールはこの機能に特化することにした。

製作したツールの入力インタフェースを図 3 に示す。本ツールは資料本文のプレーンテキストを読み込み，行単位に縦書きで表示する．ここに表示された文字をクリックすると，１文字を拡大したダイアログが表示され，色彩や形状を入力することができる．図 4 の訓点情報を入力した結果を図 3 の拡大部分に示している．単純化のために色のみではあるが，どこに点情報を入力してあるかも一目でわかるようになっている．

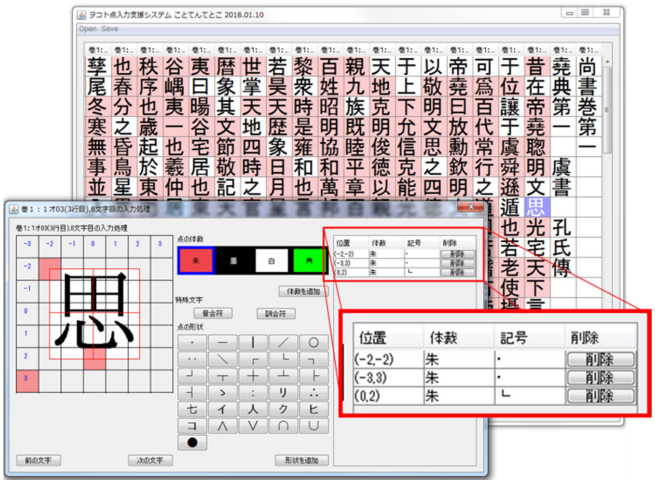

図 3　データ入力用ツール

図 4　該当する部分のヲコト点データ

また，『尚書（古活字版）』の訓点情報の体裁は主に朱色であるが，合符は墨で書かれていることが多い．本ツールは一つ前に入力した体裁を維持するようになっているが，墨の合符を入力するたびに，再び体裁を朱に戻すことを忘れることも多く不便であった．そこで，1 クリックで位置や体裁を指定しなくても合符を入力できる機能を追加した．

本ツールが出力するデータの例を図 5 に示す．他のツールでの再活用が容易なように，自由度は高いが単純な構造である json フォーマットとし，文字単位に情報を付与していくことにした．この入力されたデータには，点が句読点であるのかヲコト点であるのかなどの区別はつけていない．分析作業において，点の解釈などを検討できるようにするため，位置と体裁，形状のみを残している．

```
{
  "place" : { "lineNumber" : 3,"columnNumber" : 8 },
  "character" : "思",
  "linename" : "巻１ : １オ03",
  "elements" : [
        {"position" : {"x" : -2.0,"y" : -2.0},
        "style" : "朱",
        "mark" : "・" },
        {・・・}
  ]
}
```

図 5　生成された json データ

5.　おわりに

現在，『尚書（古活字版）』9 巻の訓点の初期入力を終え，データ校正の段階である．後日，訓点の計量結果を報告することができるだろう．

ヲコト点に関する計量の観点としては，以下のものが想定される．

1. 記号形状
2. 記号位置
3. 漢字と記号との組み合わせ
4. 1 漢字に複数の記号が使われるパターン

ⓒ 2018 Information Processing Society of Japan

また，巻 7 から巻 9 までの 3 巻の入力に要した時間は 44 時間であった．ツールの使い勝手の検証を行い，機能を向上させることが課題である．上述の計量の観点で上げた事柄は，ツールの機能向上にも寄与するものであろう．

謝辞 本研究は JSPS 科研費 17K18506 の助成を受けたものである．また，本研究は，人間文化研究機構広領域連携基幹研究プロジェクト「異分野融合による総合書物学」の国語研ユニット「表記情報と書誌形態情報を加えた日本語歴史コーパスの精緻化」による成果の一部である．

参考文献

[1] 大矢透, 仮名遣及仮名字体沿革史料, 国定教科書共同販売所, 1909.
[2] 春日政治, 西大寺本金光明最勝王経古点の国語学的研究, 斯道文庫, 1942.
[3] 中田祝夫, 古点本の国語学的研究, 講談社, 1954.
[4] 築島裕, 平安時代訓点本論考, 汲古書院, 1986.
[5] http://base1.nijl.ac.jp/~nkbthdb/, (参照 2018-04-16).
[6] http://pj.ninjal.ac.jp/corpus_center/chj/, (参照 2018-04-16).
[7] 高田智和, 訓点資料の電子化について， NINJAL Project Review Vol.4 No.1, pp.36-42, 2013.
[8] 田島孝治, 堤智昭, 高田智和, ヲコト点電子化のためのデータ構造と入力支援システムの試作, 人文科学とコンピュータシンポジウム「じんもんこん 2012」論文集, pp.211-216, 2012.
[9] 堤智昭, 田島孝治, 高田智和, 点図情報入力支援ツールによるヲコト点図の電子化, 人文科学とコンピュータシンポジウム「じんもんこん 2015」論文集, pp.185-190, 2015.
[10] 林昌也, 田島孝治, 堤智昭, 高田智和, 小助川貞次, 訓点資料の加点情報計量のためのデータ構造, 人文科学とコンピュータシンポジウム「じんもんこん 2017」論文集, pp.45-52, 2017.
[11] 李丞宰, 角筆口訣의 解讀과 飜譯 1 ―初彫大藏經의 <瑜伽師地論> 巻第五와 巻第八을 中心으로―, 太學社, 2005.
[12] 朴鎭浩, 文字生活史の観点から見た口訣, 文学第 12 巻第 3 号, pp.169-181, 2011.
[13] 堤智昭, 田島孝治, 小助川貞次, 高田智和, 訓点資料の構造化記述方式と計算機を用いた基礎計量, 情報処理学会論文誌 vol.59 No.2, pp.278-287, 2018
[14] http://dglb01.ninjal.ac.jp/ninjaldl/bunken.php?title=syousyo, (参照 2018-04-16).

ⓒ 2018 Information Processing Society of Japan