

共同研究プロジェクト紹介 萌芽・発掘型：訓 点資料の構造化記述 訓点資料の電子化について

著者	高田 智和
雑誌名	国語研プロジェクトレビュー
巻	4
号	1
ページ	36-42
発行年	2013-06
URL	http://doi.org/10.15084/00000729

訓点資料の電子化について

Sharing Classical Chinese Documents with Reading Marks in Electronic Form

高田 智和 (TAKADA Tomokazu)

1. はじめに

日本語史の研究資料の一つに、訓点資料と総称される一群がある。漢籍、仏典、国書などの漢文本文に対して、日本語で理解するための訓点を加えたもので、奈良時代から近代に至るまで、各時代・各学派（博士家や仏教流派）の特色ある訓点が施されている。これまで、日本語史研究では、文字（万葉仮名から片仮名への変遷など）、音韻（音便や日本漢字音など）、語彙・語法（漢文訓読語彙など）といった面で、訓点資料が活用されてきた。

しかしながら、日本語史研究に活用されるような訓点資料の多くは、文化財であるため、原本に接することが難しい。そのため、重要典籍については、所蔵者や研究者によって、积文（解読文）や、影印・複製が製作され、研究のための資料として提供されてきた。しかし、积文は純然たる一次資料ではなく、影印・複製はそれ自体が稀覯品化していたり、原資料に使われた朱による書き込みの情報が失われた白黒印刷であったりと、研究利用にあたって少なからず問題もある。また、訓点資料研究においては、原本のデジタル画像化や、积文の電子テキスト化など、総じてデジタル技術の導入が、他の人文学分野に比べて立ち遅れている現状もある。

そこで、本研究課題「訓点資料の構造化記述」では、研究者間で研究資料を相互に利活用できるように、原本のデジタル画像の提供と、积文の構造化記述法の提案を行う。

2. 原本のデジタル画像化と公開

近年のデジタルアーカイブは、古典籍の原本画像や電子化テキストを提供してくれるが、訓点資料の公開はまだ少なく、訓点資料研究においてはその恩恵に十分に浴しているとは必ずしも言い難い。資料のデジタル画像だけに着目しても、訓点解読のために必要な情報を保持した画像を提供しているものは少ない。デジタル画像の中に何が写し込まれていればいいのか、つまり、訓点解読のために必要な情報とは何なのかを、訓点資料の研究者自身が発してこなかったことにも、その一因が求められるだろう。

平安・鎌倉時代の訓点資料は、現在の高等学校漢文教科書にあるような、句読点、返り点、送り仮名を主体とした訓点とは違い、漢文本文に対して、句読点、科段点、ヲコト点（図1、図2参照）、声点、語順符、合符、仮名注（和訓、字音点など）、漢文注（類音注、反切、注

積など)などを加えている。特に、ヲコト点は、現代の漢文訓読では使われていない記号であり、平安・鎌倉時代の訓点資料を特徴づけている。

ヲコト点は、漢字の字画の隅に記す記号で、形状(星点、線点、鉤点など)と位置によって、対応する音節(あるいは音節列)が異なる。また、同じ形状、同じ位置にある点であっても、ヲコト点の種類が異なれば、対応する音節も異なる。例えば、字画の左隅の星点は、喜多院点では「ヲ」を表し、古紀伝点では「テ」となる。中田祝夫(1954)では、点図集に記載された代表的なヲコト点を8群に分類している。

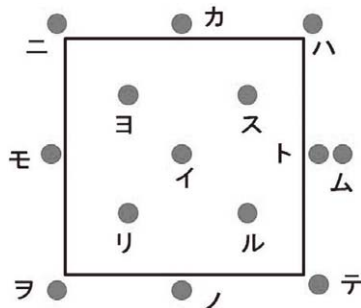


図1 喜多院点(星点)

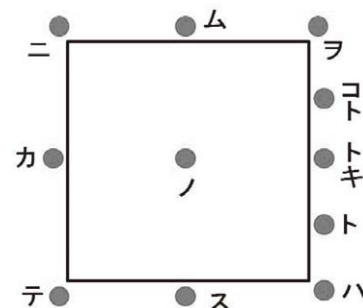


図2 古紀伝点(星点)

また、平安・鎌倉時代の訓点資料では、朱筆、墨筆、白筆などが用いられ、色彩を持っている。訓点の色彩は、加点年代や加点者の違いなどを反映している場合が多い。朱の濃い薄い、あるいは、墨の太い細いが、同一の資料で併存していることもあり、これらも加点年代や加点者の違いなどによって考えられる。

朱筆、墨筆は、蛍光灯下でも視認でき、かつ、特殊な方法を用いずにデジタル撮影ができる。これに対して白筆は、資料によっては、蛍光灯下では視認しにくい(見えない)こともある(自然光がよいとされる)。最近では、LEDライトの使用が白筆の観察に効果を発揮しているが(調査場所によってはLEDライトの使用が許可されない場合もある)、白筆が剥落していたり、薄くなっていたりする資料では、視認が難しい場合もある。同一資料の中でも、視認しやすい箇所と視認しにくい箇所とがあり、個別事例的で必ずしも一様ではない。なお、視認しにくい白点資料は、概して、通常の方法ではデジタル画像化が困難な資料である。

視認しにくい訓点の筆頭は、角筆点である。角筆点は、角筆(木や象牙などで作られた先のとがった筆記具)を用いて紙に凹みをつけて、記号や文字を記した訓点である。紙の凹みであるから、光の当たり具合如何で、視認できたり、視認できなくなったりする。また、紙の凹みであるから、それが訓点であるか(意味のある凹みであるか)否かは、研究者の慎重な判断を要する。なお、角筆点の視認と撮影には、「角筆スコープ」(斜光装置)の利用が有効である(小林芳規・吉沢康和(1991))。

このように、訓点の形状や色彩(と紙の凹み)が可読できる状態のデジタル画像を作らなければ、訓点の解読には活用できない。また、最終的には、典籍の本文を読み下すのであるから、訓点のある数文字単位の局所的な画像よりも、数行から一紙くらいまでの単位で、文

脈がたどれる程度の分量をカバーした画像であることが望ましい。

以下に、白点を含む資料のデジタル撮影の事例を紹介する。国立国語研究所蔵『金剛頂一切如来真実撰大乘現証大教王経』（以下、国語研本金剛頂経）は、平安末期移点の白点資料（喜多院点）である。国語研本金剛頂経は文化財指定品ではないため、原本へのアクセスという観点では「見やすい資料」である。

国語研本金剛頂経の撮影では、通常撮影と斜光撮影の2種類の撮影方法を用いた。通常の方法で高精細デジタル撮影を行ったところ、白点と虫損との区別がつきにくい画像となった（図3参照）。そのため、斜光撮影で白点をとばし、虫損を強調した画像もつくった（図4参照）。かつて、白黒の影印本によって、虫損を点の類いと誤認したようなこともあったため、点と虫損や汚れを視認しわけることができる水準の画像も必要である。一通りの撮影方法による画像のみで、訓点解読のすべての要求を満たそうとすることを、むしろ放棄すべきであろう。

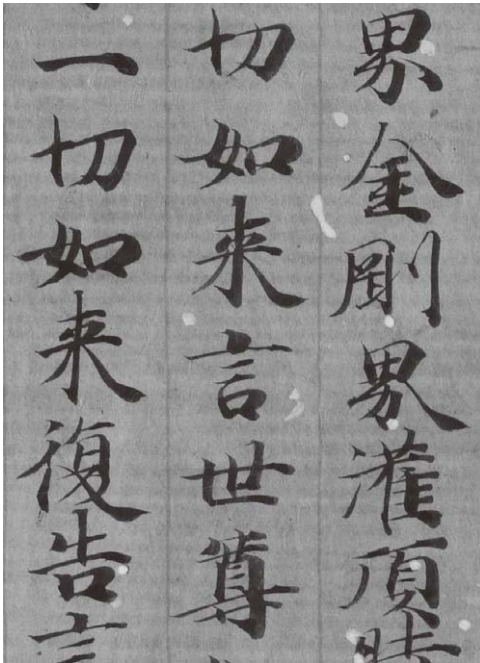


図3 通常撮影による画像

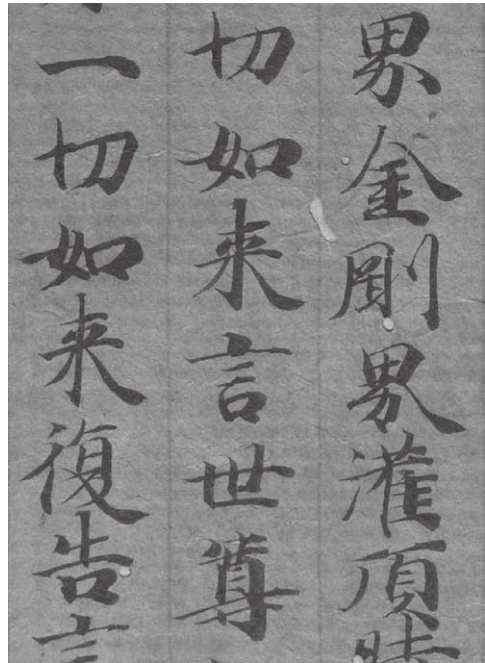


図4 斜光撮影による画像

現在、国語研本金剛頂経は、国立国語研究所ウェブサイトから公開している（<http://www2.ninjal.ac.jp/kongochokyō/>）。画像表示にはContentsViewを用い、画像を自由にスクロール、拡大縮小でき、また、通常撮影画像と斜光撮影画像を並べたり重ねて表示させたりすることもできる。

国語研本金剛頂経は、代表的な白点資料ではない。しかし、代表的な白点資料を材料として、実験的な撮影を繰り返し、インターネットでの公開まで行うことは、権利関係や技術的

制約による困難を伴う。まずは、筆者の所属機関の蔵書を材料に、撮影から画像公開までのモデルケースを示した上で、重要典籍の所蔵機関に理解と協力を求めていくのがよいのではないかと考える次第である。

3. 釈文の構造化記述方法の提案

訓点資料の基礎研究（解説）の流れは、大まかに、(1) 原本書誌調査、(2) 解説結果の記述、(3) 解説結果の共有（公表）の3段階である。

(1) 原本書誌調査は、料紙や装丁など書誌学的見地から、訓点資料の年代や資料的価値を見極める。(2) 解説結果の記述は、解説作業の中核である。さらに、ここでは、(2-1) 移点、(2-2) 釈文制作、(2-3) 書き下し文（いわゆる訓読文）制作の3段階がある。

(2-1) 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して移点本を作成する（図5参照）。そして、移点本をもとに釈文を制作する。(2-2) 釈文は、漢文本文の体裁を保ったまま翻字し、ヨ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一種の翻刻文である（図6参照）。初期解説の成果と呼べるものである。なお、図5は、台湾国家図書館蔵『史記夏本紀』の移点本の一部、図6は、小助川貞次ほか（2009）による釈文の一部である。

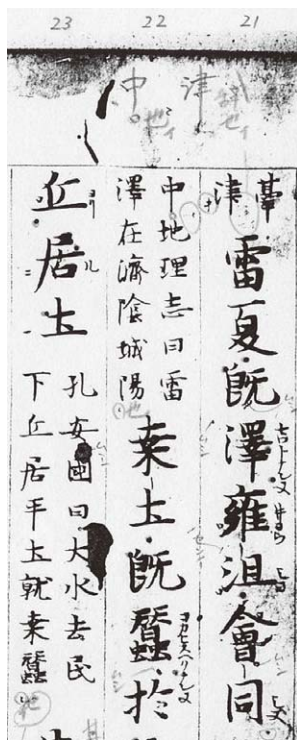


図5 移点本

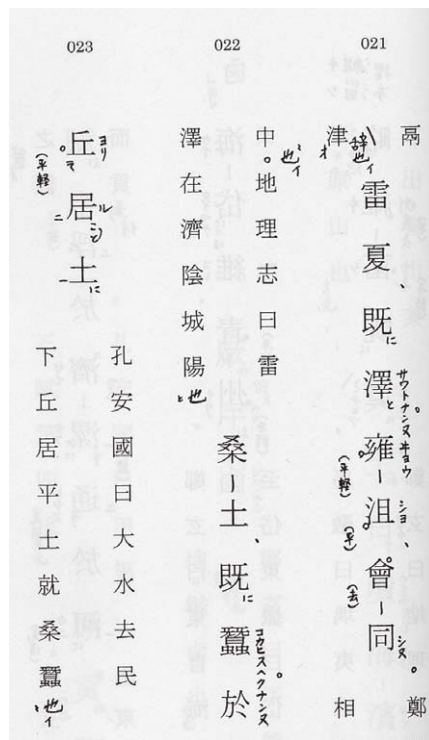


図6 釈文

移点本,あるいは, 釈文から, 書き下し文を制作する。(2-3) 書き下し文は日本語文であるから, 日本語としての文レベルでの文法・語法研究に寄与できる可能性が広がる。しかし, 前段階の釈文であっても, 従来の日本語史研究で行われてきた, 文字, 音韻, 語彙・語法など, 文字あるいは語レベルでの研究利用は可能である。

```

<!DOCTYPE 本文 [
<!ELEMENT 本文 ( タイトル, 本行 +, 注釈 + ) >
<!ELEMENT タイトル
(#PCDATA,missingCharacter)>
<!ELEMENT 本行
(#PCDATA,missingCharacter, 行, 単字, 接続,
句点, 読点, 科段点, 鉤点, 双行注)*>
<!ELEMENT 双行注
(#PCDATA,missingCharacter, 行, 単字, 接続,
句点, 読点)*>
<!ELEMENT 注釈
(#PCDATA,missingCharacter, 仮名注, 漢文注)
*>
<!ELEMENT 行 EMPTY>
<!ELEMENT 単字
(#PCDATA,missingCharacter, フコト点, 返点,
声点, 人名, 青紙)*>
<!ELEMENT 接続
(#PCDATA,missingCharacter, 単字, 合符,
人名)*>
<!ELEMENT 句点 EMPTY>
<!ELEMENT 読点 EMPTY>
<!ELEMENT 科段点 EMPTY>
<!ELEMENT 鉤点 EMPTY>
<!ELEMENT 仮名注 ( 和訓, 字音)*>
<!ELEMENT 漢文注 ( 反切, 引用)*>
<!ELEMENT 和訓 (#PCDATA, 単字, 合点)*>
<!ELEMENT 字音 (#PCDATA, 単字, 合点)*>
<!ELEMENT 反切
(#PCDATA,missingCharacter, 単字, 接続,
合点)*>
<!ELEMENT 引用
(#PCDATA,missingCharacter, 単字, 接続,
合点)*>
<!ELEMENT フコト点 EMPTY>
<!ELEMENT 返点 EMPTY>
<!ELEMENT 声点 EMPTY>
<!ELEMENT 人名 EMPTY>
<!ATTLIST 声点 声調 CDATA #REQUIRED>
<!ATTLIST 人名 class CDATA #REQUIRED>
<!ATTLIST 人名 形状 ( 星 | 線 ) #REQUIRED>
<!ATTLIST 合符 class CDATA #REQUIRED>
<!ATTLIST 合符 位置 ( 中央 | 左 ) #REQUIRED>
<!ELEMENT 合符 EMPTY>
<!ELEMENT 合点 EMPTY>
<!ELEMENT missingCharacter #PCDATA>
<!ELEMENT 青紙 EMPTY>
<!ATTLIST 行 番号 CDATA #REQUIRED>
<!ATTLIST 単字 id CDATA #REQUIRED>
<!ATTLIST 単字 属性
( 補入 | 消 | 見消 : 補入 | 見消 : 消 ) >
<!ATTLIST 接続 id CDATA #REQUIRED>
<!ATTLIST 句点 class CDATA #REQUIRED>
<!ATTLIST 読点 class CDATA #REQUIRED>
<!ATTLIST 科段点 class CDATA
#REQUIRED>
<!ATTLIST 鉤点 class CDATA #REQUIRED>
<!ATTLIST 仮名注 id CDATA #REQUIRED>
<!ATTLIST 漢文注 id CDATA #REQUIRED>
<!ATTLIST 和訓 class CDATA #REQUIRED>
<!ATTLIST 和訓 位置 CDATA #REQUIRED>
<!ATTLIST 字音 class CDATA #REQUIRED>
<!ATTLIST 字音 位置 CDATA #REQUIRED>
<!ATTLIST 半切 class CDATA #REQUIRED>
<!ATTLIST 半切 位置 CDATA #REQUIRED>
<!ATTLIST 引用 class CDATA #REQUIRED>
<!ATTLIST 引用 位置 CDATA #REQUIRED>
<!ATTLIST フコト点 class CDATA
#REQUIRED>
<!ATTLIST フコト点 形状 ( 星 | 線 | 鉤 )
#REQUIRED>
<!ATTLIST フコト点 音節 CDATA
#REQUIRED>
<!ATTLIST 返点 class CDATA #REQUIRED>
<!ATTLIST 返点 形状 CDATA #REQUIRED>
<!ATTLIST 返点 音節 CDATA>
<!ATTLIST 声点 class CDATA #REQUIRED>
<!ATTLIST 声点 形状 ( 星 | 圏 ) #REQUIRED>
<!ATTLIST 合点 class CDATA #REQUIRED>
<!ATTLIST 合点 位置 ( 右 | 左 ) #REQUIRED>
<!ATTLIST missingCharacter unicode
CDATA>
<!ATTLIST missingCharacter daikanwa
CDATA>
]

```

さて、訓点は、それ自体が漢文本文に対するアノテーションであると考えられる。釈文は、原本のレイアウトを重視した解読結果であるが、訓点の計量的な分析を行おうとすれば、句読点、科段点、ヲコト点、声点、語順符、合符、仮名注（和訓、字音点など）、漢文注（類音注、反切、注釈など）など、すべての訓点を含めた構造化ルールを設けて記述することが有益である。そこで、XMLを用いた訓点の記述方法を検討してみた。前ページに書き出したのが現段階での構造化ルールである。

4. おわりに

本研究課題は、資料研究における共同利用を推進するための、いわばインフラ整備の意味合いが強いものであった。こういった活動を通して、物理的なモノである資料そのもの、そこに書かれた言語との関係が見えてくることもあるため、今後機会を改めて継続していきたいと考えている。

●参考文献●

- 小林芳規・吉沢康和(1991)「角筆スコープの開発」『しにか』2(8):64-71.
 小助川貞次・池田証寿・渡辺さゆり・高田智和(2009)「国家図書館(台北)所蔵本史記夏本紀釈文」『訓点語と訓点資料』122:43-129.
 中田祝夫(1954)『古点本の国語学的研究』東京：講談社。

《要旨》 訓点資料原本のデジタル画像化と、釈文の電子テキスト化（構造化記述）について述べる。研究期間において、国立国語研究所所蔵の訓点資料を専用のビューアを用いて公開し、XMLによる構造化記述の方法を検討した。

Abstract: This paper describes the scanned images and digital texts used in the analysis of classical Chinese materials glossed with reading marks. During the project period, we made the images of the materials at NINJAL publicly available and proposed an XML description method for structuring the transcribed text of such glossed materials.

高田 智和 (たかだ・ともかず)

国立国語研究所理論・構造研究系准教授。博士（文学）（北海道大学）。国立国語研究所研究開発部門研究員を経て2009年9月より現職。

主な著書・論文：『例解辞典』（新版監修，ぎょうせい，2010），『電子化辞書とねじれの漢字』（『計量国語学』23(5)，2002），『漢字処理と『大字典』（『訓点語と訓点資料』109，2002），『漢字字体規範データベース』（共著，『日本語の研究』1(4)，2005）。

受賞：標準化貢献賞（日本規格協会，2007），標準化貢献賞（情報処理学会情報規格調査会，2010）。

社会活動：日本語学会電子情報委員長，計量国語学会理事，情報処理学会人文科学とコンピュータ研究会運営委員。

萌芽・発掘型共同研究プロジェクト「訓点資料の構造化記述」

プロジェクトリーダー 高田智和

(国立国語研究所 理論・構造研究系 准教授)

プロジェクトの概要

漢文訓点資料は、文字、音韻、語彙、語法などの面で、日本語史研究の資料として活用されてきた。訓点資料は歴史的・文化財的・教育的価値の高いものが多く、原本調査の難しいものが多い。そのため、重要典籍については、研究者による釈文や、影印、複製が公刊されているものもあるが、釈文は純然たる一次資料ではなく、影印、複製それ自体が稀覯品であったり、白黒印刷であったりと、研究利用にあたって少なからず問題もある。また、訓点資料研究においては、釈文の電子テキスト化や、原本の画像化など、総じてデジタル技術の導入が、他の分野に比べて立ち遅れている現状である。そこで本研究では、訓点資料の原本調査に基づいて、解読結果である釈文の構造化記述の方法を検討し、釈文と原本デジタル画像とを対照表示できるシステム開発の基礎研究を行う。