

訓点資料構造化の試み

著者	高田 智和, 田島 孝治
雑誌名	訓点資料の構造化記述 成果報告書
ページ	55-70
発行年	2013-03-29
シリーズ	国立国語研究所共同研究報告 ; 12-08
URL	http://doi.org/10.15084/00002647

訓点資料構造化の試み

高田 智和 田島 孝治

1 はじめに

平安・鎌倉時代の訓点資料は、漢文による本行・割注、書写者や校訂者等による本文訂正、加点者によるヲコト点（字画の隅に記された符号）・声点・仮名点など持ち、複雑で重層的な文書形式である。従来、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（漢文訓読語など）といった面で、訓点資料が活用されてきた。伝存する訓点資料は、平安時代だけでも 5,000 点を超えと言われる、その大部分は仏典資料である。訓点資料は、日本語史研究に活用されてきたが、これは一側面にすぎず、仏教学、歴史学、文学など、文献資料を扱う分野において共有されるものである。

近年は、パソコンの普及に伴い、各分野で文献資料の電子化やデータベース化が進み、学界の資産として共有し、研究者個人が研究や教育で活用することが行われている。しかし、言語研究資料として訓点資料を扱う分野では、対象資料の文書形式が複雑なこともあってか、漢文本文と訓点を併せての電子化やデータベース化はほとんど行われていない現状である[1]。

本稿は、訓点資料の解読を、原本から実際に行なおうとする立場で、訓点資料の基礎研究の段階において制作する釈文に対して、XML を用いた構造化記述を検討するものである。

2 訓点資料の基礎研究の流れ

訓点資料の基礎研究（解読）の流れは、大まかに次の 3 段階である。

1. 原本書誌調査
2. 解読結果の記述
3. 解読結果の共有（公表）

第 1 段階は原本書誌調査である。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。しかし、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、原本書誌調査のための閲覧に困難が伴う場合が多い。重要典籍については、複製や影印が公刊されているものがあるが、複製や影印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。博物館などで、訓点のある典籍がカラーのデジタル画像で公開されていても、ヲコト点などを確認できる画像の精度でないこともある。この場合、訓点資料の解読が公開の目的に含まれていないのであろうから、やむを得ないことである。また、角筆による訓点の有無は、特殊撮影を施さないか

1) 金木敏(1999)は、成果公表を念頭に置いた組版のマークアップである。

ぎり、原本でしか把握できないことである。

第2段階は解読結果の記述で、解読作業の中核である。ここでは、移点、釈文制作、書き下し文（いわゆる訓読文）制作の3段階がある。

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（図1参照）。

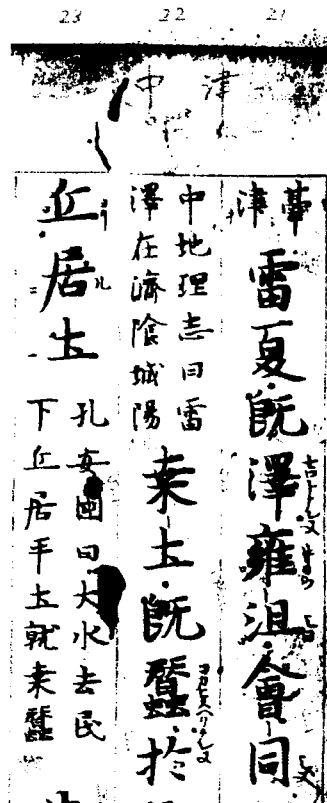


図1 移点本[2]

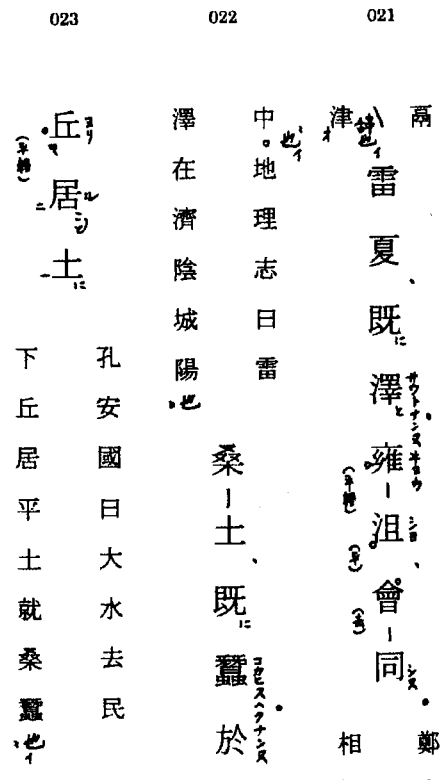


図2 釈文[3]

移点本をもとに釈文を制作する（図2参照）。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。

移点本、あるいは、釈文から、書き下し文を制作する。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。釈文でも、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。

したがって、訓点資料の基礎研究の最終段階にあたる解読結果の共有（公表）は、釈文

2) 国家図書館（台北）蔵『史記』巻第二夏本紀の一部。

3) 小助川貞次ほか(2009)より。

か書き下し文で行われる。公表の方法は、紙媒体の雑誌や図書が主で、電子化されたものは流通していない。

また、原本が一般に公開されていない場合や、海外の図書館等に所蔵されている場合は、釈文、あるいは、書き下し文によるしかなく、検証の可能性が狭められるといった問題も生じる。

3 釈文と書き下し文

現代の高等学校の漢文教科書などでは、白文に施された訓点を、ルールにしたがって読み下せば、一通りの書き下し文になる。これは、一通りの書き下し文になるように、訓法が整備された結果である。しかし、平安・鎌倉時代の訓点資料では、必ずしも書き下し文が一通りになるとは限らない。このことを、京都国立博物館蔵岩崎本『日本書紀』巻第二十二に引かれた十七条憲法の冒頭の文「以和爲貴」で説明しよう。

岩崎本の当該個所では、「和」に対して複数の訓点が加点されている（図3参照）。

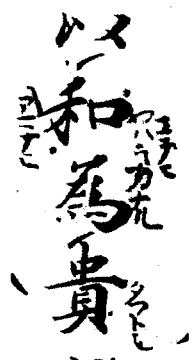


図3 複製本[4]

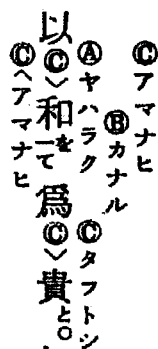


図4 釈文[5]

岩崎本の加点は4段階があるとされ、築島裕・石塚晴通（1978）による釈文（図4参照）では、訓点の色、墨の濃さ、筆跡などを原本によって調査し、平安中期末の加点をA、院政期の加点をB、室町時代宝徳3（1451）年及び文明6（1474）年の加点（一条兼良加点）をCと、加点年代を比定している。これにしたがえば、当該個所は次の三通りの書き下し文がつくられるであろう。

- A（平安中期末）：和（ヤハラク）を以て貴と為。
- B（院政期）：和（ヤハラカナル）を以て貴と為。
- C（室町時代）：和（アマナヒ）を以て貴（タフトシ）と為。

書き下し文が一通りにならないのは、加点が複数回行われて、訓点に層があるという岩

4) 『復刻日本古典文学館 日本書紀 巻第二十二 推古』（日本古典文学会、1972）より。

5) 築島裕・石塚晴通(1978)より。

崎本の特性が大きな要因である[6]。

訓については、「和」の解釈が、「ヤハラク」→「ヤハラカナル[7]」→「アマナヒ」と、加点点年によって変わったこと、「貴」は室町時代に「タフトシ」と確実に訓んでいることがわかる。平安時代に「貴」をどう訓んだか、あるいは「以」「為」の訓は、当該個所から確証は得られない。

返読については、室町時代の加点点には、「以和」「為貴」にそれぞれ雁点があるので、語順転倒を行なったことがわかる。平安時代の加点点では、「和」に返点を兼ねた「て」をあらわすヲコト点加点点されているので、「以和」は返読をしている。しかし、「為貴」に返読の確証は見られない。したがって、A、Bの書き下し文は、次のA'、B'のようにあらわすこともできる。

A'〈平安中期末〉：和（ヤハラク）を以て為貴と。

B'〈院政期〉：和（ヤハラカナル）を以て為貴と。

ヲコト点については、平安中期末の加点点を、院政期でも室町時代でも、そのまま訓読の際に利用したであろうという推定のもと、BとCの書き下し文を作成している。つまり、同一の本に複数の加点点者がいる場合、後の加点点者は前の加点点を訓み継いで、解釈を変えるべきところは新たに加点点するという前提を設けて、ここでは解読を行なったのである。

一般に、訓点資料の書き下し文では、解読者の解釈が入り込んだ結果、一通りにならないことも、可能性として起こりうる。特に、訓点がない箇所解読者が補読を試みるような場合である。築島裕・石塚晴通（1978）が、加点点年代を比定しているにもかかわらず、公表したものが積文であるのは、解読者の解釈によって解読結果に影響が大きく生じる書き下し文を避けたことも、一因として考えられる。

前出の小助川貞次ほか（2009）による国家図書館（台北）蔵『史記』巻第二夏本紀の解読では、加点点の層を大きく3段階（宝治2（1248）年安倍時貞加点点、建長8（1256）年安倍為貞加点点、文和3（1354）年惟宗守俊加点点）に比定したが、公表した積文では、それぞれの訓点に加点点年代を記述しなかった。理由は、すべての訓点の加点点年代を判別しきれていないからである。したがって、夏本紀全巻にわたって、加点点の3段階に対応する書き下し文三通りをつくることは、現状ではまだできないのである。

4 訓点資料の構造

訓点資料は、それ自体、構造を持っていると見なすことができる。大きくは、本行と双行注からなる漢文本文と、それらに対する注釈にあたる訓点ということになる。白藤禮幸（1990）は、漢文訓読を加点点者の注釈活動として扱うことを提案しており、これにした

6) 訓点資料の訓点の層については、十七条憲法と同じ例を用いて、時空間情報の観点から富山日出夫・高田智和（2007）で扱ったことがある。

7) 院政期点は、平安中期末の「ヤハラク」の「ク」の上から、「カナル」を書き足しているため、「ヤハラカナル」を意図していると考えられる。

がえば、句読点、返点、ヲコト点、声点、仮名点など、読み下しに直接関わる加点者の書き込みのほか、読み下しのための解釈に関わる注釈書や字書・音義の引用や、漢文本文の校訂、対校による異本注記など、注釈レベルのすべての書き込みを広く訓点として扱うことができるであろう。

また、訓点の物理的な形状や形態は、訓点が変わらず意味や役割を担っている。前述のように、訓点の層は加点年代や加点者の別を反映し、これは、訓点の色（朱点、墨点、白点など）や筆記具の別（毛筆か角筆か）によって知られる。

符号の形状にも、星点、圏点、線点、鉤点などがあり、ヲコト点／声点の弁別などに機能している。

同じ形状の符号であっても、付される位置によって意味が異なる。例えば、国家図書館（台北）蔵史記夏本紀の朱点のヲコト点は、次のようである（図5参照）。星点が付された位置が、変わらず音節を決定する。左下の場合、字画の隅に接近していれば「て」であるが、字画からやや離れていれば、返点を兼ねた「て」となる。しかし、どのくらい離れていたら返点になるのか、明確な基準はないようであり、解読者が文脈から判断することになる[8]。

被注字、被注字句に対する仮名点や漢文注の位置も、訓点の層を反映する場合がある。一般に、最初の加点では右側、次の加点では左側、書き込むスペースがなくなれば、欄外の頭注や、裏書になると考えられる。

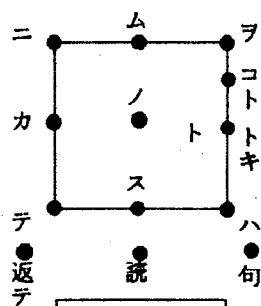


図5 ヲコト点図

符号が変わらずか、文字（言語）が変わらずかも、訓点様式の時代変遷と関わってくる。初期のヲコト点（符号）主体のものから、時代が下るにつれて仮名点主体のものへと移行していく。仮名点も語形の一部（語頭あるいは語末の音節）を示すものから、語形全体を表示する、現代の振り仮名に類するものへと変わっていく。漢字音の表記も、反切や類音注から、仮名点の多用へと移っていくようである。

釈文は、原本のレイアウトを重視した解読結果であるから、訓点の物理的な形状や形態を考慮した構造化記述を考える必要がある。

8) 返点を兼ねた「て」は、返る先の文字を特定する符号を伴わない。そのため、読み順が特定できる雁点や二点などとは、やや性質が異なる。

5 XMLによる構造化記述—京都国立博物館蔵『世説新書』巻第六残巻を例に—

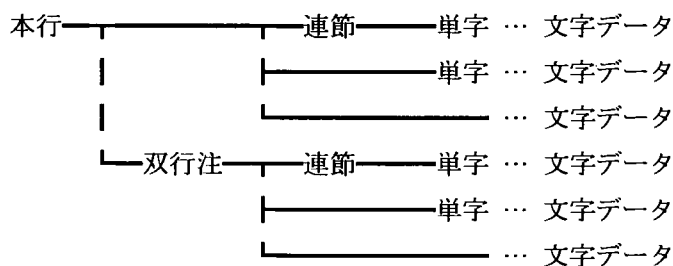
訓点資料の構造化記述に際して、XMLを試してみることにした。近代以降の日本語コーパスにおいて実績があること、目的は異なるが、漢文訓点のマークアップに前例があること、歴史学の文献資料の構造化においても実績があることなどの理由による。

今回、XMLによって構造化を試みる資料は、京都国立博物館蔵『世説新書』巻第六残巻である。漢文本文は7世紀末の唐写本、10世紀初頭頃加点のヲコト点主体の訓点資料である。訓点は朱点が2種、角筆点が1種の計3種であり、加点は3段階と考えられる。『世説新書』巻第六残巻のようなヲコト点主体のよりシンプルな構造の訓点資料でまず記述方法の骨格をつくり、その上で、京都国立博物館蔵岩崎本『日本書紀』巻第二十二や、国家図書館（台北）蔵『史記』巻第二夏本紀のような、仮名点や漢文注を豊富にもつ複雑な訓点資料に拡張していく意図から、本資料を選択した。

『世説新書』巻第六残巻の冒頭10行について、筆者による本文翻字を以下に示す。JIS外字は=で表わして丸括弧でUnicodeと大漢和辞典番号を添え、割注は[]、補入は〈 〉で括っている。

- 1 : =(U+90D7,M39413)太尉晩節好談既雅非所經而
- 2 : 甚矜之 [中興書曰鑒少好學博覽羣書學雖不章句而多所通綜也]
- 3 : 後朝觀以王丞相末年多可恨每
- 4 : 見必欲苦相規誠王公知其意每
- 5 : 引作他言臨當還鎮故命駕詣丞
- 6 : 相翹鬚 =(U+53B2,M03041)色上坐便言方當永別
- 7 : 必欲言其所見意滿口重辭殊不
- 8 : 溜王公攝其次日後面未期亦欲
- 9 : 盡所懷願公勿復談 =(U+90D7,M39413)遂大〈隕氷〉矜而
- 10 : 出不得一言

今回は、訓点が施される単字や文字連節が属する漢文本文の層と、単字や文字連節に施される訓点が仮名点や漢文注といった文字（言語）で示される注釈の層と、2層を分けて構造を設けてみた。漢文本文の層の入れ子関係は、次のとおりである。



ヲコト点、返点、声点など単字に付される符号は単字の要素とし、合符など文字連節に付される符号は連節の要素とする。また、句点、読点などは、文字データの階層の要素とみなす。また、漢文本文での漢字の物理的な位置を示すため、行番号の要素を文字データの階層に設ける。

訓点の要素では、次のように属性を記述する。『世説新書』巻第六残巻の冒頭 10 行で使
用したものを例示する。class 属性には、訓点の層を記述する。この資料では、3 種の層そ
れぞれの加点年代が明確ではないので、訓点の色と筆記具を用いることにした。ヲコト点
については、ここでは、字画に付された位置ではなく、音節に解釈して記すことにしたが、
より精緻な記述を行うためには、韓国口訣資料の解読で利用されている 5×5 分割などの導
入も有効であろう。

```
<ヲコト点 class="朱" 形状="星" 音節="に" />
<返点 class="朱" 形状="星" 音節="て" />
<合符 class="朱" 位置="中央" />
<人名 class="薄朱" 形状="星" />
<句点 class="角" 形状="星" />
<読点 class="朱" 形状="星" />
```

単字、文字連節に仮名点や漢文注が付されている場合には、ID を設け、注釈を参照させ
る。注釈の層の入れ子関係は、次のとおりである。

```
注釈 ─── 仮名注 ─── 和訓 … 文字データ
      |                └── 字音 … 文字データ
      └── 漢文注 ─── 反切 … 文字データ
                    └── 引用 … 文字データ
```

仮名注（主として省画仮名、万葉仮名で書かれたもの）には和訓と字音、漢文注（主と
して漢字、漢文で書かれたもの）には反切と引用の要素をそれぞれ設ける。『世説新書』巻
第六残巻の冒頭 10 行に出現した要素は和訓のみである。例示する。

```
<注釈>
  <仮名注 id="t041">
    <和訓 class="朱" 位置="右">云</和訓></仮名注>
  <仮名注 id="t048">
    <和訓 class="朱" 位置="右">モ</和訓></仮名注>
</注釈>
```


和訓にさらに声点がつくような場合が想定されるため、和訓、字音、反切、引用の要素の下位にも、単字や文字連節の要素を置くことを考えていたが、『世説新書』巻第六残巻にそのような事例はなかった。

『世説新書』巻第六残巻の第 9 行目に見られる本行への補入「臆氷」は、「臆」「氷」のそれぞれ一文字ずつを単字要素にし、属性に「補入」を記述することで処理した。文字列の補入を、文字列単位ではなく単字単位で扱ったのは、補入の 2 文字目の「氷」と、それに続く「矜」とが合符で結ばれ、訓読上の一単位を形成しているため、XML の入れ子のねじれを回避しようとしたからである。なお、青紙は、補入と関係していると考えられるが、独立した要素として扱った。

```
<単字 id="t044" 属性="補入">臆
  <ヲコト点 class="朱" 形状="星" 音節="て" /></単字>
<接続 id="r002">
  <単字 id="t045" 属性="補入">氷</単字>
  <単字 id="t046">矜<青紙 /></単字></接続>
```

『世説新書』巻第六残巻は、JIS X 0208 文字セットを用いて翻字したため、JIS 外漢字の処理が必要になる。外字は=であらわし、タグで表現することにした。属性には、Unicode と大漢和辞典親字番号を記述する。

```
<missingCharacter unicode="U+90D7" 大漢和="M39413">=</missingCharacter>
```

ここまで述べたルールによって、『世説新書』巻第六残巻の第 1 行目と第 2 行目を構造化すると、以下のようになる (XML 宣言等は省略)。

<本文>

```
<タイトル>京都国立博物館蔵『世説新書』巻第六残巻</タイトル>
```

```
<本行>
```

```
<行 番号="1" />
```

```
<missingCharacter unicode="U+90D7" 大漢和="M39413">=</missingCharacter>
```

```
太尉
```

```
<読点 class="薄朱" 形状="星" />
```

```
晩
```

```
<単字 id="t001">節
```

```
<ヲコト点 class="朱" 形状="星" 音節="に" />
```

```
<ヲコト点 class="角" 形状="星" 音節="に" /></単字>
```

```
好
```

<単字 id="t002">談

<ヲコト点 class="朱" 形状="星" 音節="を" />

<ヲコト点 class="角" 形状="星" 音節="を" />

<返点 class="朱" 形状="星" 音節="て" />

<返点 class="角" 形状="星" 音節="て" /></単字>

既

<単字 id="t003">雅

<ヲコト点 class="朱" 形状="線" 音節="より" /></単字>

非

<単字 id="t004">所

<ヲコト点 class="朱" 形状="星" 音節="に" />

<ヲコト点 class="角" 形状="星" 音節="に" /></単字>

<単字 id="t005">經

<ヲコト点 class="朱" 形状="鉤" 音節="たる" /></単字>

<単字 id="t006">而

<ヲコト点 class="朱" 形状="星" 音節="を" />

<ヲコト点 class="角" 形状="星" 音節="を" /></単字>

<行 番号="2" />

甚

<単字 id="t007">矜

<ヲコト点 class="朱" 形状="鉤" 音節="たる" /></単字>

之

<句点 class="薄朱" 形状="星" />

<双行注>

中興書曰

<単字 id="t008">鑿

<人名 class="薄朱" 形状="星" /></単字>

<単字 id="t009">少

<返点 class="朱" 形状="星" 音節="て" />

<返点 class="角" 形状="星" 音節="て" /></単字>

好學

<句点 class="朱" 形状="星" />

<句点 class="角" 形状="星" />

博覽羣

<単字 id="t010">書

<ヲコト点 class="朱" 形状="星" 音節="を" /></単字>

<句点 class="朱" 形状="星" />

學

<読点 class="朱" 形状="星" />

雖不章句

<読点 class="朱" 形状="星" />

<読点 class="角" 形状="星" />

而

<読点 class="角" 形状="星" />

多所通綜也

</双行注>

画面出力用に、XML から HTML に変換したデータを、Web ブラウザで表示した釈文の画面例を示す (図 6 参照)。訓点の表記は築島裕・石塚晴通(1978)に準じ、画面出力の利点を活かして、朱点と角筆点で一致するものを青、朱点のみのものを赤、角筆点のみのものを緑に色分けしている。冒頭 10 行だけを見ても、朱点と角筆点の一致率が極めて高く、同時に、朱点のみの訓点が相当数あり、角筆点のみの訓点が極めて少ないことから、角筆点をなぞるようにして朱点を精緻に加えたと見るべきであろうか。

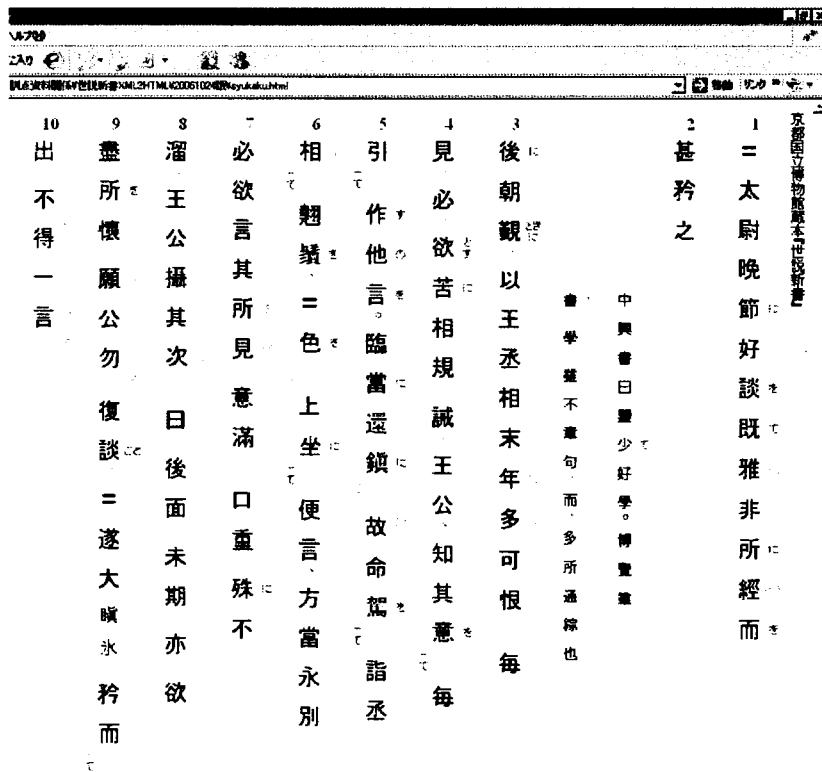


図 6 画面出力例

6 XMLによる構造化記述—京都国立博物館蔵岩崎本『日本書紀』巻第二十二の一部を例に—

『世説新書』巻第六残巻の構造化記述にあたって設計したタグは、次の23種である。

本文

タイトル、本行、双行注、単字、連接、注釈、仮名注、漢文注、和訓、字音反切、引用

行、句点、読点、missingCharacter

ヲコト点、返点、声点、人名、合符、青紙

これらのタグを用いて、京都国立博物館蔵岩崎本『日本書紀』巻第二十二の第103～104行目（十七条憲法の冒頭を含む推古天皇12年夏の冒頭）について構造化を試みると、科段点、鉤点、合点の3種のタグが追加される。

構造化例を以下に記す（XML宣言等は省略）。

<本文>

<タイトル>岩崎本日本書紀巻第二十二推古紀</タイトル>

<本行>

<行 番号="103" />

<科段点 class="朱" />

夏四

<単字 id="t001">月

<ヲコト点 class="朱" 形状="星" 音節="の" /></単字>

丙

<単字 id="t002">寅

<ヲコト点 class="朱" 形状="星" 音節="の" /></単字>

朔戊

<読点 class="朱" 形状="星" />

辰

<連接 id="r001">皇太子

<合符 class="墨 C" 位置="中央" /></連接>

<単字 id="t003">親</単字>

<単字 id="t004">肇

<ヲコト点 class="朱" 形状="星" 音節="て" /></単字>

<単字 id="t005">作

<返点 class="墨 C" 形状="二" /></単字>

<行 番号="104" />

<連接 id="r002">憲法
 <合符 class="墨 C" 位置="左" /></連接>
 <連接 id="r003">十七
 <単字 id="t006">條
 <返点 class="墨 C" 形状="一" /></単字></連接>
 <句点 class="朱" />
 <鉤点 class="朱" />
 <単字 id="t007">一
 <ヲコト点 class="朱" 形状="星" 音節="に" /></単字>
 <単字 id="t008">日</単字>
 <読点 class="朱" />
 <単字 id="t009">以
 <返点 class="墨 C" 形状="レ" /></単字>
 <単字 id="t010">和
 <ヲコト点 class="朱" 形状="星" 音節="を" />
 <返点 class="朱" 形状="星" 音節="て" /></単字>
 <単字 id="t011">爲
 <返点 class="墨 C" 形状="レ" /></単字>
 <単字 id="t012">貴
 <ヲコト点 class="朱" 形状="星" 音節="と" /></単字>
 <句点 class="朱" />
 </本行>
 <注釈>
 <仮名注 id="r001">
 <和訓 class="墨 C" 位置="右">ヒツキノミコ</和訓></仮名注>
 <仮名注 id="t003">
 <和訓 class="朱 A" 位置="右">ミ</和訓></仮名注>
 <仮名注 id="t004">
 <和訓 class="墨 C" 位置="右">ハ</和訓></仮名注>
 <仮名注 id="t005">
 <和訓 class="墨 B" 位置="右">ツクリキ</和訓>
 <和訓 class="墨 C" 位置="右下">タマフ</和訓></仮名注>
 <仮名注 id="r002">
 <和訓 class="朱 A" 位置="右">イツクシキ</和訓>
 <和訓 class="墨 C" 位置="右下">(イツクシキ)ノリ</和訓></仮名注>
 <仮名注 id="r003">
 <和訓 class="朱 A" 位置="右">トヲチアマリナ、ヲチ</和訓></仮名注>

```

<仮名注 id="t006">
  <和訓 class="墨 B" 位置="右">ヲチ</和訓></仮名注>
<仮名注 id="t008">
  <和訓 class="墨 C" 位置="右">ク</和訓></仮名注>
<仮名注 id="t010">
  <和訓 class="朱 A" 位置="右">ヤハラク</和訓>
  <和訓 class="墨 B" 位置="右">(ヤハラ) カナル</和訓>
  <和訓 class="墨 C" 位置="右">アマナヒ</和訓>
  <和訓 class="墨 C" 位置="左">
    <合点 class="墨" 位置="右" />アマナヒ</和訓></仮名注>
<仮名注 id="t012">
  <和訓 class="墨 C" 位置="右">タフトシ</和訓></仮名注>
</注釈>
</本文>

```

7 おわりに

京都国立博物館蔵『世説新書』巻第六残巻は、ヲコト点主体の訓点資料であるため、構造化のためのルールが少なく済んだが、今後、京都国立博物館蔵岩崎本『日本書紀』巻第二十二や国家図書館（台北）蔵『史記』巻第二夏本紀など、仮名点や漢文注による注釈が豊富な資料に対して、本格的に適応させた場合には、ルールの見直しや、根本的な変更が発生することも起こりえよう。本稿は、試行錯誤の始まりを述べたにすぎない。

釈文は、訓点資料の解説成果の一つではあるが、一次資料ではない。研究利用にあたっては検証の必要があり、原本は難しくとも、原本の画像データと並べて、併せて活用することが望ましい。岡本隆明（2008）などで提案されているような、原本画像を参照できる「デジタル釈紋」のシステムが構築されれば、訓点資料の研究や高等教育に益するところ大であろう。

師茂樹（2007）は、訓点資料を含めた漢文のコーパスの必要性を述べている。訓点資料を扱う研究者も活用できるようなコーパスを目指そうとすれば、試行錯誤の時間がしばらく必要である。今後の課題としたい。

最後に、今回設定した DTD を示す。

```

<!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, 連接, 句点, 読点, 科段点, 鈎点, 双行注)*>
  <!ELEMENT 双行注 (#PCDATA, missingCharacter, 行, 単字, 連接, 句点, 読点)*>

```

<!ELEMENT 注釈 (#PCDATA,missingCharacter, 仮名注, 漢文注)*>
 <!ELEMENT 行 EMPTY>
 <!ELEMENT 単字 (#PCDATA,missingCharacter, ヲコト点, 返点, 声点, 人名, 青紙)*>
 <!ELEMENT 連接 (#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, 連接, 合点)*>
 <!ELEMENT 引用 (#PCDATA,missingCharacter, 単字, 連接, 合点)*>
 <!ELEMENT ヲコト点 EMPTY>
 <!ELEMENT 返点 EMPTY>
 <!ELEMENT 声点 EMPTY>
 <!ELEMENT 人名 EMPTY>
 <!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 連接 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|圈) #REQUIRED>
 <!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>
 <!ATTLIST 合点 class CDATA #REQUIRED>
 <!ATTLIST 合点 位置 (右|左) #REQUIRED>
 <!ATTLIST missingCharacter unicode CDATA>
 <!ATTLIST missingCharacter daikanwa CDATA>
]

参考文献

- 岡本隆明 (2008) 「古文書・典籍を対象とした文字管理システムとその可能性」、『情報処理学会研究報告』2008-CH-078、pp.77-84
 金水敏 (1999) 「kuntent2e.sty について」、<http://www.let.osaka-u.ac.jp/~kinsui/tex/top.htm>.
 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、pp.43-129
 後藤真 (2007) 「日本の漢文史料とマークアップ—新たなデータベースの可能性に向けて—」、『漢字文献情報処理研究』8、pp.83-92
 白藤禮幸 (1990) 「注釈としての漢文訓読」、『国語と国文学』67-2、pp.1-19
 田中牧郎 (2005) 「言語資料としての雑誌『太陽』の考察と『太陽コーパス』の設計」、『雑誌『太陽』による確立期現代語の研究—『太陽コーパス』研究論文集—』、pp.1-48、国立国語研究所
 築島裕・石塚晴通 (1978) 『岩崎本日本書紀』、貴重本刊行会
 築島裕 (1986) 『平安時代訓点本論考』、汲古書院.

- 當山日出夫・高田智和 (2007) 「文献研究における時空間情報の構造についての考察—東洋古典籍研究の視点から『日本書紀』を事例として—」、『人文科学とコンピュータシンポジウム論文集 デジタルアーカイブ—デジタルアーカイブと時空間の視点—』、pp.31-38
- 師茂樹 (2007) 「漢文のマークアップ—現状と課題—」、『漢字文献情報処理研究』8、pp.70-72
- 山口昌也・高田智和・北村雅則・間淵洋子・小林正行・西部みちる (2008) 『『現代日本語書き言葉均衡コーパス』における電子化フォーマット ver2.0 (国立国語研究所内部報告 LR-CCG-08-01)』、国立国語研究所
- 山崎直樹 (2007) 「訓点付き漢文の返り点から統語情報を導出し XML で構造化する試み」『漢字文献情報処理研究』8、pp.73-82