

# 訓点資料の書き下し文自動生成を目的とした ヲコト点を中心とする訓点の計量分析

田島 孝治<sup>1,a)</sup> 堤 智昭<sup>2</sup> 高田 智和<sup>3</sup>

受付日 2019年5月16日, 採録日 2019年11月7日

**概要:** 本論文では, 訓点資料の書き下し文を自動生成することを目的とし, 書き下し文を自動生成するうえで問題となる, 1つの漢字に複数のヲコト点が付与されている場合の分析を行い, ヲコト点の読み順が定まるかを検証した. ヲコト点は記号訓点の一種で, 漢字に付された位置と形状により読みは定まるが, ヲコト点が複数あった場合, 順番は定まらない. たとえば「シ」, 「テ」, 「ニ」と3つのヲコト点がついていた場合に「ニシテ」と順を定めて書き下すためには, 読み順の定義が必要である. 対象とした資料(国立国語研究所蔵『尚書(古活字版)』)には121種類, のべ672個のヲコト点の組合せが登場するが, いずれも読み順が一意に定まることが確認できた. また, 漢字の読み仮名の一部を表す場合のみに使われるヲコト点も抽出することができた.

キーワード: 訓点資料, 漢文訓読, ヲコト点, 書き下し文の自動生成

## Quantitative Analysis of *Kunten* Materials for Interpreted Text Generation

KOJI TAJIMA<sup>1,a)</sup> TOMOAKI TSUTSUMI<sup>2</sup> TOMOKAZU TAKADA<sup>3</sup>

Received: May 16, 2019, Accepted: November 7, 2019

**Abstract:** This paper describes quantitative analysis of the gloss in the *Kunten* Materials. It aims at an automatic interpreted text generation. For a description the *Kunten* Materials, we developed the *Iten* tool and constructed a data of gloss in the “Shangshu” (old type print version). This constructed data is suitable for the quantitative analysis with computer. We analyzed the reading order of the multi wokototen marks on a character using the data. Define the reading order of wokototen marks is important to generate the interpreted text. For example, a character has 3 kinds of wokoten these shows “シ”, “テ”, “ニ”. In this case, the reading order is “ニシテ”. We need to identify the general patterns that can appear in the documents, and define the reading order for these patterns. We checked every combination pattern of the wokototen marks in this document. There are 121 combination patterns, and we fixed the order. In addition, we extract the wokototen marks that show how to read the Chinese character.

**Keywords:** *Kunten* Material, wokototen, automatic generation of the vernacular reading text

### 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ててきた. 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.

<sup>1</sup> 岐阜工業高等専門学校  
National Institute of Technology, Gifu College, Motosu, Gifu  
501-0495, Japan

<sup>2</sup> 筑波大学  
University of Tsukuba, Tsukuba, Ibaraki 305-8577, Japan

<sup>3</sup> 国立国語研究所  
National Institute for Japanese Language and Linguistics,  
Tachikawa, Tokyo 190-8561, Japan

a) ktajima@gifu-nct.ac.jp

漢文訓読のために書き入れられた句読点、語順点、声点、ヲコト点などの記号や仮名を訓点といい、訓点を記入した典籍・文書を訓点資料という。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・語法などの史の変遷を解明するために、訓点資料が活用されてきた。特に、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、訓点資料が重視されている。

訓点資料による主要な研究成果には、大矢透 [1]、春日政治 [2]、中田祝夫 [3]、築島裕 [4] などがあり、特に春日政治 [2] は訓点に従って漢文を日本語文に直した書き下し文を提供し、現在の一般的な訓点資料研究の成果公表方法となっている。しかし、訓点資料研究へのデジタル技術導入研究自体が少なく(岡本隆明 [5]、田中勝ほか [6] などがある)、着手が始まった段階といえる。なかでも、訓点研究の成果として一般的な書き下し文を自動生成することは、現状では困難である。

訓点資料では、現代の漢文訓読では使用されないヲコト点が多用され、これが書き下し文作成を困難にする一因である。ヲコト点は、助詞や助動詞、漢字の読みを表す記号で、漢字に付される位置と形状によって読みが決定される。また、ヲコト点の種類は宗派や学派によって異なり、資料ごとの個別事例もあるため、書き下し文作成には、ヲコト点に対する知識と分析を蓄積する必要がある。

筆者らは、これまで、ヲコト点を中心に、訓点資料の電子化に取り組んできた(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]、朴鎮浩 [12] など)を日本のヲコト点に援用し、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])。しかし、点図集収載のヲコト点は訓点資料の中で使われる可能性があるヲコト点の一覧であり、いわば文字コード表のようなものである。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と同様に、実際の訓点資料に記入された訓点を扱わなければ、訓点の使用傾向を知ることができない。そこで、実際の訓点資料の訓点を電子化することで、ヲコト点をはじめとする訓点の計量分析を行うことが次の課題となる。

本論文では、訓点資料解読の一般的な成果提示方法である書き下し文の自動生成を目的として、実際の訓点資料の訓点を電子化したうえで計量分析を行う。今回は特に、書き下し文を自動生成するうえで問題となる、1つの漢字に複数のヲコト点が付与されている場合の分析を行い、ヲコト点の読み順が定まるかを検証する。

## 2. 書き下し文作成までの流れ

訓点資料の解読では、移点、積文作成を経てから、書き下し文作成に至る。

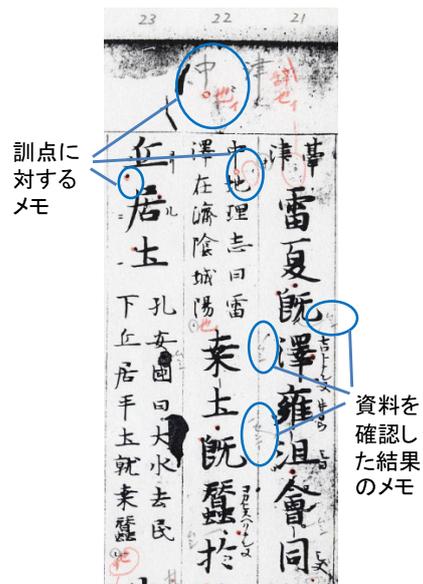


図 1 訓点資料と移点結果のメモ

Fig. 1 The results of transferring gloss.

### (1) 移点

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、移点は以下のような手順で行われる。

- ① 本文と割注を転記した用紙を用意する。
- ② 未記入の点図を用意する。
- ③ 本文の基本情報を確認する。
- ④ 本文を読みながら訓点を転記する。

移点では、用意した本文に訓点を移していくと同時に、点図を作成する。図 1 のように本文とともに付与された墨や朱による書き込みを写し取っていく。また、用紙のサイズ、丁数、奥書など基本的な書誌に関する情報を確認、記録する。

### (2) 積文作成

移点結果をもとに積文を作成する。積文の例を図 2 に示す。積文は、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である。初期解読の成果と呼べるものである。図 2 の積文では、句点を「.」、読点を「,」、仮名点を片仮名、ヲコト点を平仮名に置き換えて記述している。声点は「○」を漢字の周囲に記入し、声調を括弧書きで添えている。

### (3) 書き下し文作成

移点結果、あるいは積文から書き下し文を作成する。原漢文の語順を日本語語順に直し、仮名点やヲコト点で示された助詞・助動詞や送り仮名を埋め込んだ漢字仮名交じり文が一般的な書き下し文である。たとえば、図 1 および図 2 の「居土」は、一二点に従って語順を転倒させ、「居」「土」に付されたヲコト点をそれぞれ平仮名「こ」「に」に置き換え、「居」の仮名点「ル」を送り仮名にし、「土に

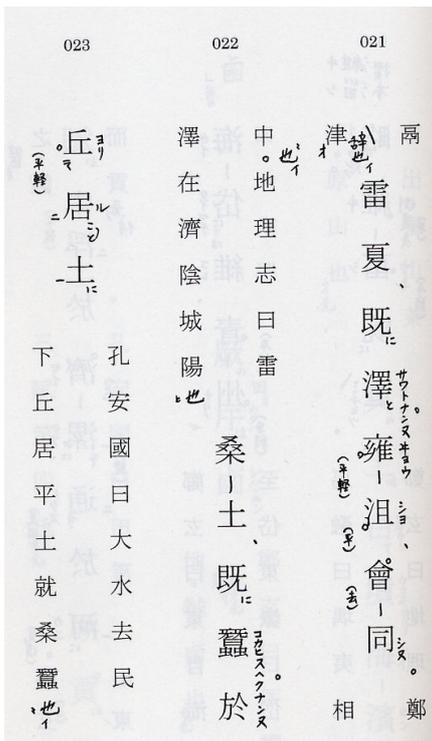


図 2 釈文

Fig. 2 The printed re-transcription of glossed text.

居ルこと」と書き下し文を作成する。日本語としての文レベルでの文法・語法研究に利用するならば、このような日本語語順に直した書き下し文が求められる。

このように、訓点資料の書き下し文作成では、移点結果があることが必須である。そのため、書き下し文自動生成を目指した基礎分析向けのデータ作成においては、移点作業そのものを支援するツールと、移点結果をデータ化する記述方式の整備が必要となる。

### 3. 対象資料とデータ作成

#### 3.1 対象資料

今回、電子化および分析の対象とする資料は、国立国語研究所蔵『尚書（古活字版）』春本（第1冊，巻1-3）である [15]。この資料は慶長年間（1596-1615）刊のものであり、川瀬一馬 [16] に『尚書（古活字版）』の第3種本と言及されている。冬本（第4冊，巻10-13）と刊記・奥書が欠けているが、伏原（清原）宣條（1720-1791）の書入本と推定される。全編にわたりヲコト点，語順点（一二点やレ点），仮名点，注釈が朱と墨により記入され，「宣賢」の書入れがあることから，江戸時代の伏原（清原）家当主宣條が，架蔵本を用いて清原宣賢（1475-1550）をはじめとする伝来の訓点を書き写したのと考えられる。明経道の博士家である清原家の『尚書』の訓法を伝えるものとして，貴重な訓点資料である。

資料は専用のviewerに加え，ページごとの画像データがjpegファイルで公開されているため，細部まで拡大して確

表 1 電子化した訓点の種類

Table 1 Target glosses on the classical Chinese texts.

訓点の名称	訓点の持つ意味・解釈
科段点	段落
句読点	文・文中の区切り
音合符/訓合符	語のまとまりと語の読み方（音読み，訓読み）を表す
音読符/訓読符	語の読み方（音読み，訓読み）を表す
声点	漢字のアクセント（声調）を表す
ヲコト点	助詞・助動詞，語形の一部（読み仮名や送り仮名に相当）を表す
校符	誤写や補入・削除などの本文訂正，異本との対校を表す

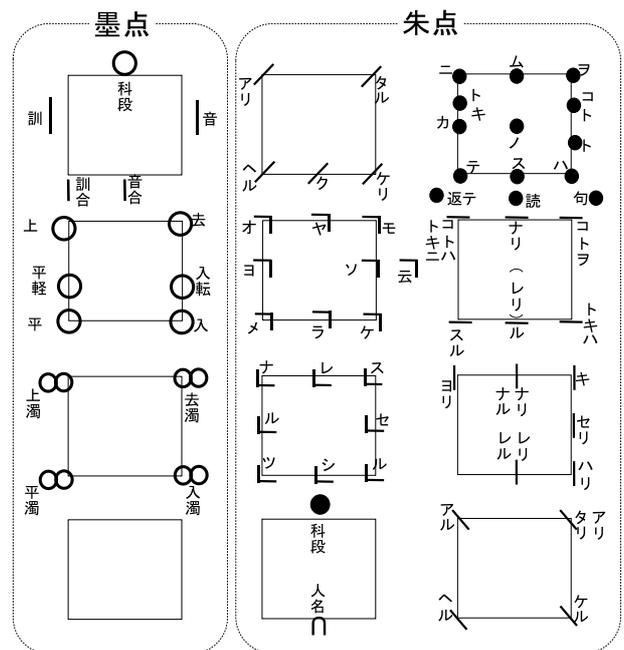


図 3 『尚書（古活字版）』の点図

Fig. 3 The target glosses on the document.

認することができる。冊子本であるため，画像データは1丁に対し表裏が存在し，半丁あたり8行構成である。

#### 3.2 電子化の対象とした訓点

今回電子化の対象とした訓点は，科段点，句読点，音合符/訓合符，音読符/訓読符，声点，ヲコト点，校符の記号訓点7種である。表1に電子化した訓点の種類と機能をまとめる。校符は使い方によって位置が異なる。この資料では，文字の左側に「○」を付け，誤字の修正や，異本で使われている別の字形の文字を欄外に記入する，あるいは文字と文字の間に「○」を付けて，補入を表す。

記号訓点のうちヲコト点は，書き下し文を作成する際には仮名に置き換えて，漢字とともに記述する。電子化の対象とした記号訓点に対して作成した点図を図3に示す。点図を作ることで，形状が同じ「・」であるヲコト点と句読点の位置による区別が明確となり，ヲコト点の仮名への置き換えも一意に行うことができるようになる。しかし記述



図 4 複数のヲコト点が記された場合の例

Fig. 4 An example of the glosses in the document.

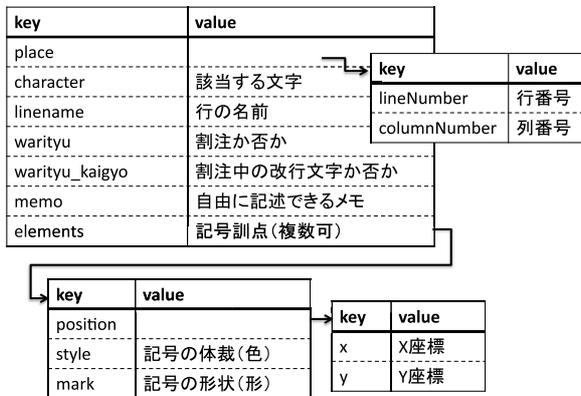


図 5 JSON データの構造

Fig. 5 The JSON format for glosses.

の順は一意に定まらない。たとえば図 4 の例では、「色」には左上の「・」で表せる「ニ」、左下の「・」で表される「テ」、中央下の「L」で表される「シ」と 3 つのヲコト点が付与されている。専門家であれば、ヲコト点の組合せから「色にして」と書き下すことができるが、これを自動的に行うには、組合せのパターンを調査し、並び順を定義する必要がある。記号訓点を電子化対象とすることで、資料中の漢字 1 字に対するヲコト点の組合せの種類を調べ、ヲコト点が付与されている漢字の種類によらず書き下しの順が定まるのかを調査することができ、書き下し文の自動生成におけるヲコト点の読み順を定めることが可能になる。

### 3.3 データの記述方式

電子化の対象とした訓点を記述するためのデータ構造を図 5 に示す。また、図 4 の文字に関する電子化結果を図 6 に示す。書き下し文の生成やデータベースの構築に利用することを考えると、別のプログラミング言語でも利用可能なデータ構造が望ましい。そこで、自由度は高いが単純な構造である JSON フォーマットを採用した。このデータ構造では資料の本文の 1 文字を単位に主となる表を作り、記

```
{
  "place": {
    "lineNumber": 555,
    "columnNumber": 1
  },
  "character": "色",
  "linename": "巻 6 : 1 ヲ05",
  "warityu": false,
  "warityu_kaiqyo": false,
  "elements": [ {
    "position": { "x": 0.0, "y": 2.0 },
    "style": "朱", "mark": "・"
  }, {
    "position": { "x": -2.0, "y": -2.0 },
    "style": "朱", "mark": "・"
  }, {
    "position": { "x": -2.0, "y": 2.0 },
    "style": "朱", "mark": "・"
  } ]
}
```

図 6 JSON データの記述例

Fig. 6 An example of JSON data for the glosses.

号訓点は“elements”として複数記述する。

このデータ構造では、入力作業時に、記号の形状や記号の位置（以後、記号形状、記号位置と記す）、朱や墨など体裁を記録するだけでよいようにするため、句読点やヲコト点など訓点の種類では区別しない。この単純な入力方法により訓点資料に精通していないユーザであってもデータの入力作業が可能である。一方でこれらのデータは、専門家がデータ校正や分析を行い、ヲコト点や文意の解釈を検討するためには十分である。なお、点の位置を表す座標表現については高田 [14] で詳しく述べたとおり、文字中心を原点とした二次元の座標系とする。この X, Y 座標の値は整数で -3~3 の範囲である。

### 3.4 データ入力用ツール

訓点情報の入力手法の 1 つとしてソフトウェアを開発し、訓点の電子化に用いた。このソフトウェアを移点ツールと呼ぶ。今回このツールを開発したのは、JSON データはプログラミング言語では扱いやすいが、人間が直接テキストを入力していくには煩雑であるためである。

製作したツールの入力インタフェースを図 7 に示す。本ツールは訓点資料の本文をプレーンテキスト化したデータを読み込み、行ごと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、訓点情報を入力することができる。入力後、色彩・形状を選択してから、文字上の該当箇所をクリックする。形状や色は必要に応じて追加できる。入力後は表示されたマスの色が変わることのできる位置に情報を入力したかが分かるようになっている。

今回のツールは表 1 の記号訓点 7 種を入力できるようになっている。また、文字に対してメモを残す機能を設け、虫食いによる穴なのか区別がつかない、異本注記にのみヲコト点がついているなどの入力時に気づいたことを記録で

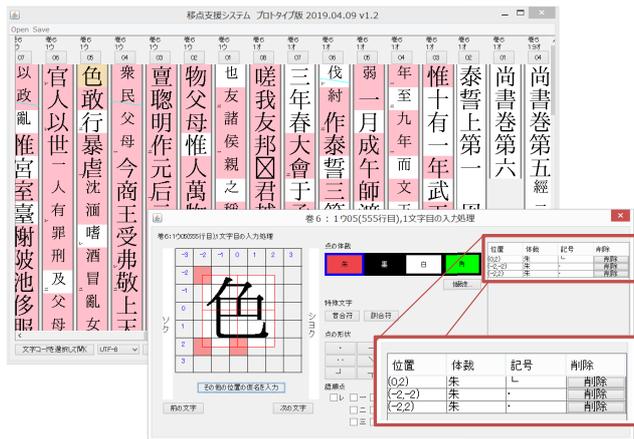


図 7 データ入力用ツール  
Fig. 7 The main interface of developed tool.

きる。語順点、仮名点、合点などの入力は今後の開発目標であり、改良を続けていく予定である。

### 3.5 データの入力と校正

電子化は①データ入力、②入力結果の確認、③統計処理を利用した校正、④複数ヲコト点の記述順の決定の4段階で実施した。①と②の作業は同一であり、本文画像と点図を見ながら移点ツールを用いて、1文字ごとに記号訓点の訓点情報を入力する。②の作業も実施する作業は変わらない。ただし、すでに入力された情報が表示されるため、入力間違いや、入力漏れを発見した場合は修正・追加する。③の作業は訓点を記号形状、体裁、記号位置ごとに集計し、度数が10未満のものは入力ミスの可能性があると考え、本文画像と入力データを見合わせて確認・修正する。④の作業は、ヲコト点が複数記入された箇所をすべて抽出し、本文画像と入力データを見合わせ、ヲコト点の位置と、その読み順を確認する。

①データ入力と②入力結果の確認は、それぞれ担当者を変えて岐阜高専の学生が実施した。両名とも計算機が使える、訓点知識のない学生である。③統計処理および④複数ヲコト点の記述順の決定は筆者らが行った。

②~④の工程では、①で入力したデータの間違いを校正した。②の作業においては追加、削除など421文字に対する修正を行った。③の作業により40文字に対する修正を実施した。さらに、この作業において、文字の左側か文字と文字の間に記述された○の記号が見つかった。これらは声点と同じ記号形状ではあるが、校符であった。このため、声点の位置にない○の記号がある112カ所に関しては、すべて確認し校符とそうでないものに再度分類した。この結果、99個の校符が新たに発見された。④の作業においては、18文字の訓点が適切な読みに対応していなかったため修正した。

表 2 資料に付与された訓点の度数

Table 2 Statistics of the glosses in the document.

訓点の名称	度数	
	墨点	朱点
科段点	4	102
句読点	0	2,752
音読符/訓読符	518	0
音合符/訓合符	2,876	0
声点	217	0
ヲコト点	0	8,459
校符	99	6
小計	3,714	11,319
合計		15,033

表 3 墨点の記号形状と度数

Table 3 The shape and frequency of the black glosses.

記号	度数	全墨点数に対する割合	割合の累計
	3,401	89.64%	89.64%
○	274	7.22%	96.86%
∞	83	2.19%	99.05%
∟	23	0.61%	99.66%
その他	13	0.34%	100.00%
合計	3,794	100.0%	-

## 4. 計量結果

### 4.1 資料の文字数と訓点の総数

資料の基礎データを測るために、資料の文字数および、表1に示した訓点の個数を計測した。資料の総文字数は16,605文字であった。また、表1に示す訓点を種類別に集計した度数を表2にまとめる。集計は点図に合わせ、墨点と朱点に分けてまとめた。その結果、記号訓点は句読点、合符、ヲコト点が大割合を占めていることが分かった。さらに、ほとんどの訓点は墨点と朱点のいずれかのみが使われており、両方が使われているのはごく一部の科段点と校符に限られ、例外的であることも分かった。

### 4.2 訓点の記号形状と記号位置

墨点の記号形状の集計結果を表3にまとめる。墨点は図3の点図の記号形状、記号位置から外れているものではなく、89%が「|」点の音読符/訓読符または音合符/訓合符である。これらの記号は単語の特定に使うものの、具体的な読み方を記述してはいないため、書き下し文に直接影響を与えることはない。このため、書き下し文を機械的に生成する際には問題にならない。一方で、朱点は記号形状と記号位置により具体的な読み方を表している。たとえば、「下」という漢字には「テ」や「ス」を表す朱点が付けられることがあるが、これらは「下テ(くだりて)」と「下ス

表 4 朱点の記号形状と度数

Table 4 The shape and frequency of the vermilion glosses.

形状	度数	全朱点数に対する比率	累計比率
・	9,818	86.74%	86.74%
ㄥ	549	4.85%	91.59%
—	508	4.49%	96.08%
ㄣ	118	1.04%	97.12%
●	102	0.90%	98.02%
（	81	0.72%	98.74%
／	52	0.46%	99.20%
＼	43	0.38%	99.58%
	42	0.37%	99.95%
○	6	0.05%	100.00%
合計	11,319	100.0%	-

Y \ X	-3	-2	-1	0	1	2	3
-3	0	0	0	102	0	0	0
-2	0	1,449	0	425	0	1,544	0
-1	0	8	0	0	0	211	0
0	0	127	0	964	0	57	0
1	0	0	0	0	0	332	0
2	0	770	0	825	0	1,202	0
3	551	0	0	1,905	0	0	847

図 8 朱点の記号位置

Fig. 8 The frequency of the vermilion glosses in each position.

(くだす)」のように書き下し文でも語形として書き分けなければならない。そこで、朱点の記号形状と記号位置に関して、さらに詳しく分析する。

朱点に関しては、記号形状と記号位置ごとに集計を行った。集計結果を、表 4 と、図 8 にまとめる。なお、記号位置を調べた際に、文字の真下の「・」の「ス」と「一」の「ル」が共起するといった、同じ座標に複数の形状の異なる朱点がつけられていることがないか調べたが、このようなケースは存在しなかった。朱点の記号形状は 86% が星点「・」である。「ㄥ」「—」も一定数使われているが、星点に比べると十分の一にも満たない数である。記号位置は文字の中央下 (0, 3) が多く、文字の四隅にも多く分布している。文字の中央下にある朱点は読点のみである。また、右下 (3, 3) も句点のみであり、読点よりは少ないものの多く使われている。一方、文字の上に重ねてつけられた記号に関しては、形状と合わせて確認していく必要がある。星点の形状と記号位置を組み合わせると、(2, -2) であれば「ヲ」、(-2, -2) であれば「ニ」、(2, 2) であれば「ハ」、

表 5 1 漢字に複数のヲコト点がある場合の延べ数と異なり数

Table 5 The total number and the number of the kinds of the glosses combination.

1 漢字に付与されたヲコト点の個数	延べ数	異なり数
4	1	1
3	100	24
2	571	96
合計	672	121

(-2, 2) は「テ」となる。4 つ角の中で「テ」の度数がやや少ない。しかし、この点図においては、(-3, 3) も返り点兼用の「テ」を表しているため、資料中に「テ」が現れにくいというわけではない。

### 4.3 1 漢字に複数のヲコト点が付与されるパターン

仮名に置き換えた記号訓点の並び順を考えるために、1 つの漢字に複数の朱点が付与されているパターンを抽出して分析を行う。並び順を考えなければならない朱点はヲコト点のみである。ただし、点図で「返テ」と表した星点は対象としない。この「返テ」点は、返り点の機能も持つため、このヲコト点がついている文字に続けて読むことはないため、読み順に影響がないためである。

対象としたヲコト点が複数ついている文字の計量結果を表 5 に示す。該当する文字は 672 文字で、組合せの種類は 121 種類であった。最大で 4 個のヲコト点が 1 漢字についているパターンがあるが、2 個の場合が多い。

組合せごとの登場回数を数え、度数 10 以上をまとめたものを表 6 に示す。最も多く登場する組合せは「シ」と「テ」であり、登場箇所を確認し、「シテ」の順で読むことが分かった。同様にすべての組合せの登場場所を確認し、「トハ」、「ニシテ」、「ヲハ」、「ニシ」、「スルコト」というように、対象の資料中に存在した 121 種類に関しては、すべて読み順を定めることができた。

## 5. 考察

### 5.1 記号形状に対する考察

表 3, 表 4 を記号形状に注目して合計すると、「・」は朱点でのみに現れるものの最も多く 9,818、次いで墨点と一部の朱点に使われる「|」が 3,436 であった。記号訓点 15,033 個に占める割合で見ると、65.30% と 22.85% となり、記号訓点の 88% 以上がこの 2 つの記号形状であり、他の形状に比べて圧倒的に多いことが分かる。「・」形状はヲコト点であり、テ、ニ、ヲ、ハなど助詞、助動詞として多用されているためであると考えられる。一方、「|」をヲコト点として使う数は少なく、音合符/訓合符または音読符/訓読符として使われる場合がほとんどである。音合符、訓合符は、熟語を示すために資料内では多用されてい

表 6 1 漢字に複数のヲコト点がある場合の組合せと度数 (度数 10 以上のもの)

Table 6 The frequency of the glosses combination.

ヲコト点の 組み合わせ	度数	比率	累計比率
[シ], [テ]	70	10.42%	10.42%
[ト], [ハ]	65	9.67%	20.09%
[ニ], [シ], [テ]	53	7.89%	27.98%
[ヲ], [ハ]	40	5.95%	33.93%
[ニ], [シ]	40	5.95%	39.88%
[スル], [コト]	28	4.17%	44.05%
[ニ], [ス]	24	3.57%	47.62%
[コト], [ハ]	21	3.13%	50.74%
[ニ], [ハ]	17	2.53%	53.27%
[ノ], [カ]	16	2.38%	55.65%
[スル], [ニ]	14	2.08%	57.74%
[ト], [シ], [テ]	14	2.08%	59.82%
[ヲ], [テ]	13	1.93%	61.76%
[スル], [ヲ]	13	1.93%	63.69%
[スル], [トキハ]	13	1.93%	65.63%
[ニ], [テ]	12	1.79%	67.41%
その他 (105 種類)	219	32.59%	100.00%
合計	672	100.00%	-

るため、墨点では最も多く、資料に記載された記号訓点全体でも「・」に次ぐ順位になったと考えられる。

また、堤ほか (2018) [9] では、主要ヲコト点図に登場する記号形状は「・」が最も多く、次いで「ㄣ」「|」「一」の順で多く使われているとされている。一方、今回の実験結果によると、朱墨すべての記号訓点をあわせ、多く使われていた形状は「・」「|」「ㄣ」「一」の順である。ヲコト点図に記載された点は、同一の点が複数出ることがない点の集合であるため、今回の実験結果と一概に比較することはできないが、実際の訓点資料においてもヲコト点図によく出てくる形状が多く使われている傾向にあることが分かる。一方で、実際の資料では、ヲコト点図では 3 番目に多く使われているとされている「|」が、2 番目に多く使われている「ㄣ」より多く記載されているなど、ヲコト点図から分かる特性とは違う傾向がある。このことから、実際に書き下し文を作成するときには、ヲコト点図から分かる情報だけでなく、実資料を計量した結果に基づいた知見も合わせて方式を検討する必要がある。

## 5.2 記号位置に対する考察

記号位置ごとに朱点を集計した結果、1 つの記号位置に

異なる形状の訓点と同時に存在しないことが明らかになった。たとえば図 7 の点図では、「ス」と「ル」を表す朱点だが、形状は異なるが同じ座標である (0,2) を使っているような関係である。このため、この 2 つを組み合わせで「スル」もしくは「ルス」を作ることにはできない。このように、同じ座標に点が現れないという特徴を使えば、読み順ルールを作る際に検討すべきヲコト点の組合せを制限できる。このように組合せを制限できることは、今後ヲコト点の読み順のルールを拡張する際に有用である。なお、今回の点図では (-2,2) に 1 つの記号で「スル」を表す朱点が存在している。「スル」という要素はサ変動詞連体形であり、訓読において欠かせないものである。

## 5.3 1 漢字に複数の訓点が付与される場合の考察

1 漢字に複数の訓点が付与されている場合について、その組合せを確認した結果、すべての組合せは読み順が一意に定まることが確認できた。この結果をルール化したことにより、対象資料のヲコト点を適切な順で並べて書き下すことが可能になった。

組合せを確認していくなかで、書き下し文を作る際には読み仮名に相当するヲコト点が存在することに注意しなければならないことも明らかとなった。表 6 では度数が少ないためその他に含まれているが、図 9 のようなヲコト点の組合せがある。この例では、どちらも「行」という字に対し「オ」を表すヲコト点が付けられている。この「オ」は、「行」という漢字を「ギョウ」ではなく「おこなう」と読むよう指示するための読み仮名の一部であり、単純に書き下し文に含めてしまうと、読みが重複してしまう。さらに、語頭の読みであるため、そのまま文字の後ろに挿入して書き下すと、送り仮名のように見えてしまい、読む際に違和感がある。たとえば「(お) 行ことを」とするなど表記を変えて字の前に示したり、そもそも書き下し文から除去したりするなどの対応が必要であろう。

今回の対象資料中に現れた読み仮名に該当するヲコト点は、「オ」「ケ」「ナ」「モ」の 4 種類であり、いずれも読み仮名として抽出することが可能と考えられる。「オ」は「行」以外には「於 (おいて)」、「治 (おさむ)」、「興 (おこす)」、「己 (おのれ)」、「畏 (おそる)」、「起 (おこる)」、「教 (おしふ)」、「重 (おもんず)」の 8 種類の漢字と併記されるが、単独で出現する場合も含め読み仮名以外に利用することはなかった。「ケ」も同様で、読み仮名以外で登場することはなかった (「可 (べけんや)»)。「ナ」は読み仮名としても送り仮名としても使われているが、送り仮名は 1 例しかない。一方で、読み仮名は 12 例存在するが、漢字の種類は「乃 (なんぢ)」、「何 (なに)」、「生 (なす, なる)」、「汝 (なんぢ)」の 4 種類に限定されており、漢字により読み仮名かどうかを区別することができる。同様に、「モ」は読み仮名と送り仮名の場合の度数が拮抗しているが、読み仮名と

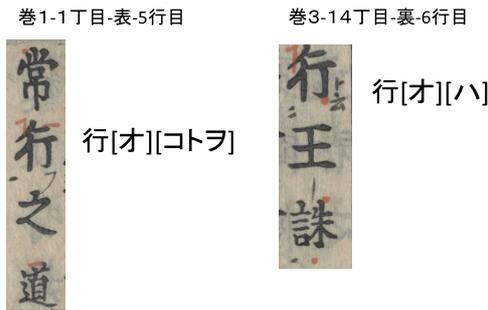


図 9 字の音を表すヲコト点の例

Fig. 9 Examples of the gloss combination that include the reading of the kanji.

して付けられた漢字は「者(も の)」「衆(も ろもろ)」「用(も ちゆ)」の3種類しかなくこちらも分類可能である。

## 6. おわりに

### 6.1 成果

本論文では、訓点資料解説の一般的な成果提示方法である書き下し文の自動生成を目的として、実際の訓点資料の訓点を電子化したうえで計量分析を行った。

訓点資料の書き下し文作成を困難にする一因となっているのは、漢文訓読では使用されないヲコト点を中心とした記号訓点であった。そこで記号訓点を電子化して記述する手法を提案し、電子化のための移点ツールを製作した。そして、国語研蔵『尚書(古活字版)』に記された記号訓点を電子化して、計量分析した。

この結果、「資料原文の語順を保持した書き下し文」の生成の際に問題となる、1漢字に複数のヲコト点が付与されていた場合の適切な並べ替えが可能であることを明らかにした。また、漢字の読み仮名を表すヲコト点の扱いについての問題点を示し、このヲコト点の種類や、付与された漢字との関係により、読み仮名とするか送り仮名とするかを区別することができるようになった。

### 6.2 今後の課題

電子化した資料を使って書き下し文を作るうえで、返り点兼用のヲコト点「て」の扱いが課題である。返り点兼用の「て」は、たとえば「明[ニ][シ]俊徳[ヲ][テ]」のように現れる。このヲコト点は「徳」という漢字についているものの、読み順としては「明[ニ][シ][テ]俊徳[ヲ]」とその漢字よりも前の漢字に続けて読むものである。返って読む場合に、他のヲコト点との間に入ることはないため、読み順は問題にならないが、どの漢字に戻して読めばよいかを定める方法が確定していない。資料を見てみると、返り点兼用の「テ」は一二点などの語順点とともに現れることが多いため、語順点の電子化により返る先を定める方法が有効であると考えている。語順点が付与されていない場合には、さらなる検討が必要である。「テ」は直前の動詞となる漢字

に付与される可能性が高いため、漢文の形態素解析を行って動詞を探す、動詞のくる可能性が高い直前の句読点の後の文字に機械的に結び付けるなどの方法が考えられ、今後検討していく必要がある。

また、今回提案した電子化と計量分析を通じた書き下し文の自動生成は、一般的な訓点資料を理解するために使われてきた主要ヲコト点図に基づいている。このため、他の漢籍、仏典を対象としても有効であると考えられる。しかし、現存する訓点資料は個性が高いものも多い。今後、他の訓点資料に本電子化手法を適用できるかの検証も実施していく必要がある。

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] 岡本隆明：コンピュータによる訓点資料の整理について，人文科学とコンピュータシンポジウム「じんもんこん 2008」論文集，pp.275-282 (2008)。
- [6] 田中 勝，村川猛彦，宇都宮啓吾：訓点資料を対象とした翻刻支援システムの構築，情報処理学会論文誌，Vol.59, No.2, pp.288-298 (2018)。
- [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 昌哉，田島孝治，堤 智昭，高田智和，小助川貞次：訓点資料の加點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 翻譯 1—初彫大藏經의 瑜伽師地論〉卷第五와 卷第八을 中心으로，太學社 (2005)。
- [12] 朴 鎮浩：文字生活史の観点から見た口訣，文学第 12 卷第 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] 高田智和：ヲコト点の座標表現，国立歴史民俗博物館研究報告，第 192 集，pp.171-181 (2014)。
- [15] 日本語史研究資料 [国立国語研究所蔵]：尚書(古活字版)，入手先 (<http://dglb01.ninjal.ac.jp/ninjaldb/bunken.php?title=syousyo>) (参照 2019-05-07)。
- [16] 川瀬一馬：古活字版之研究(増補版)，Antiquarian Booksellers Association of Japan (1967)。



田島 孝治 (正会員)

岐阜工業高等専門学校. 2007年東京農工大学大学院工学府情報コミュニケーション工学専攻博士前期課程修了. 2010年同大学院工学府電子情報工学専攻博士後期課程修了. 2011年より岐阜工業高等専門学校電気情報工学科で, 情報通信技術, 計算機技術の教育研究や, 人文学と情報学の複合的な研究に従事. 現在, 同校電気情報工学科准教授. 電子情報通信学会, 社会言語科学会, 日本語学会各会員. 博士(工学).



堤 智昭 (正会員)

筑波大学. 2010年東京農工大学工学部情報工学科卒業. 2012年同大学大学院工学府博士前期課程修了. 2015年同大学院工学府博士後期課程修了. 現在, 筑波大学人文社会学系助教. モバイルネットワークエミュレータ, 時刻情報応用システム, 自律分散型インターネットセキュリティ基盤に関する研究に従事する一方, 漢字・訓点の情報処理, 通時コーパスの構築・応用に関する研究にも従事. 日本語学会会員. 博士(工学).



高田 智和 (正会員)

国立国語研究所. 1975年生まれ. 1999年北海道大学文学部卒業. 2004年同大学大学院文学研究科博士後期課程修了. 2005年独立行政法人国立国語研究所研究員を経て, 2009年から大学共同利用機関法人人間文化研究機構国立国語研究所准教授. 漢字字体の研究, 漢字資料の文献学的研究, 漢字情報処理, 文字コード標準化に従事. 日本語学会, 計量国語学会, 訓点語学会各会員. 博士(文学).