

訓点資料の構造化記述方式と計算機を用いた基礎計量

堤 智昭^{1,a)} 田島 孝治² 小助川 貞次³ 高田 智和⁴

受付日 2017年5月9日, 採録日 2017年9月5日

概要: 本研究では, 訓点の一種であるヲコト点を対象とし, 計算機を用いて解析することを目的としたヲコト点の構造化記述方式を提案する. 提案方式に基づいてヲコト点図の電子化を支援する入力支援ツールを開発し, 主要ヲコト点 26 種を電子化する. さらにそれを用いてヲコト点の基礎計量を行う. 基礎計量では, ヲコト点を構成する要素である「読み」・「位置」・「形状」を対象とした. その結果, 最も多く登場する「読み」は「ス」「ナル」「ナリ」「タリ」であり, 形状は「・」であることや, 漢字の四隅に多くのヲコト点が付与されることなどを定量的に示した. また, 主要ヲコト点 26 種の電子化処理と基礎計量の結果を受け, 提案する構造化記述方式の有効性を確認した.

キーワード: 訓点資料, ヲコト点, 構造化記述方式

Computerized Method of KUNTEN and Quantitative Analysis of KUNTEN Using Computer

TOMOAKI TSUTSUMI^{1,a)} KOJI TAJIMA² TEIJI KOSUKEGAWA³ TOMOKAZU TAKADA⁴

Received: May 9, 2017, Accepted: September 5, 2017

Abstract: In this research, we propose the computerization method which made the WOKOTOTEN that guiding marks for rendering Chinese classics into Japanese are one kind the subject. The purpose of this method is analysis WOKOTOTEN using computer. We develop the input support tool with which computerization of a WOKOTOTENZU is supported and computerize 26 kinds of main WOKOTOTEN based on a suggestion system. Next a basic measure of a WOKOTOTEN is performed using that data. We made “reading”, “location” and “form” which is the element of the WOKOTOTEN is composed the subject in this basic measure. We got a quantitative evaluation result the “reading” on which I appear most being “SU”, “NARU”, “NARI” and “TARI” as a result. The validity of the proposed method was confirmed by computerization processing of 26 kinds of main WOKOTOTEN and a result of the basic quantitative.

Keywords: KUNTEN material, WOKOTOTEN, computerized method

1. はじめに

歴史学・文学・仏教学・文化学など様々な分野において

歴史研究を行ううえでの資料として古典籍と呼ばれる書物が活用されている。古典籍の中には、かつての東アジアの共通書記言語であった漢文を用いて記述されている書籍が多く存在する。その中には、訓点資料と総称される、漢籍、仏典、国書などの漢文に本文とは別のアノテーションがついた書籍がある。このアノテーションは訓点と呼ばれ、中国語である漢文を読者が用いる言語（日本語や韓国語）として理解し読むために、当時の読者により加えられた注釈である。日本における訓点には、奈良時代から現代に至るまで、句読点、返読点、仮名点など様々なものが存在する。その中で本研究では、平安・鎌倉時代の漢文文献に多く存

¹ 東京電機大学
Tokyo Denki University, Adachi, Tokyo 120–8551, Japan
² 岐阜工業高等専門学校
National Institute of Technology, Gifu College, Motosu, Gifu 501–0495, Japan
³ 富山大学
University of Toyama, Toyama 930–8555, Japan
⁴ 国立国語研究所
National Institute for Japanese Language and Linguistics, Tachikawa, Tokyo 190–8561, Japan
a) t_tsutsumi@mail.dendai.ac.jp

在する訓点であるヲコト点を主な研究対象とする。

漢文に訓点が付与された資料は、本文が漢文であることに加え、それに付与されている訓点が複雑な書き入れである。そのためこれらの資料を、計算機を用いて解析することが難しく、訓点をテキスト化するための記述方式も規準となる方式は存在しない。そこで本研究では、ヲコト点のついた訓点資料を、計算機を用いて扱うための環境構築を目的とした構造化記述方式を提案する。提案する構造化記述方式は、訓点そのものの研究に用いるために、読者が漢文本文と訓点情報を組み合わせて読むことによって得られる書き下し文ではなく、訓点情報そのものをテキスト化する。

本稿では、まず提案した方式に従ってヲコト点図データを作成する支援ツールを作成し、それを用いてヲコト点図の電子化を行う。さらに電子化したヲコト点図データを用いて、ヲコト点の基礎計量を行い、構造化記述方式の有効性を検証する。ヲコト点図データは、築島 [1] に記載された主要ヲコト点 26 種の情報をもとに電子化したデータを整形したものをを用いる。

2. ヲコト点

2.1 ヲコト点の概要

ヲコト点は、平安・鎌倉時代の訓点資料に多く記載されている記号の一種であり、星点、線点、鉤点といった多様な形状の記号を用い、助詞や助動詞、活用語尾などを表す。ヲコト点は漢字の字画の四隅や内部、周辺に付与され、漢字のどの位置にどの形状の点が加えられるかによって読みが異なる。ヲコト点の位置・形状と読みとの対応の種類は、時代や流派によって異なり、東大寺点や喜多院点などの名称がつけられている。現在までに確認されているヲコト点の種類は 200 種を超えている。中田 [2] は、代表的なヲコト点 26 種をヲコト点の歴史の変遷・発達を考慮し 8 つの群 (第 1 群点, 第 2 群点, 第 3 群点, 第 4 群点, 第 5 群点, 第 6 群点, 第 7 群点, 第 8 群点) に分類している。本稿では以降、これら代表的なヲコト点 26 種を主要ヲコト点 26 種と呼ぶ。

2.2 ヲコト点図

図 1 に示すようにヲコト点の形状と位置によって、それぞれがどのような読みを持つかを漢字に見立てた四角い枠に図示したものをヲコト点図 (または単に点図) と呼ぶ。この 1 つの枠を壺と呼ぶ。ヲコト点図には通常、ヲコト点の形状によって複数の壺が含まれるため、ヲコト点図は壺の集合として定義できる。先述した東大寺点や喜多院点は、壺の集合体である。ヲコト点の形状は「・」「-」「.」など多岐にわたり、一般的な計算機が表現できる文字セットには存在しない記号も使われている。

ヲコト点が付与された漢文を訓読する場合、このヲコ



図 1 ヲコト点図例 (喜多院点)。文献 [1] より引用
Fig. 1 Example of Wokototenju (Kitanoin).

ト点図に従って、付与されているヲコト点を書き下すことで、日本語として理解することが可能となる。たとえば、「東大寺点」に従ってヲコト点が付与された資料は、「東大寺点」のヲコト点図に従って書き下すことで、日本語として理解可能な文章へ変換することができる。情報工学的な視点からは、「漢文とヲコト点から構成される文章」を「日本語の語法に従って訳読可能な文章」へと変換するための符号化方式に相当するものがヲコト点図であるとすれば、ヲコト点図がどういったものが容易に理解ができるであろう。

3. 訓点資料電子化の課題

訓点資料は研究・教育に用いるため、また資料の保存を行うため様々な形での電子化が試みられている。訓点資料の電子化手法は次の 2 種類に分類できる。デジタル画像として計算機に取り込む一次資料化と、電子テキストとして取り込む二次資料化である。

一次資料化は貴重な原本資料の保存や、拡大縮小などの画像処理により、肉眼では確認しにくい資料状態の把握ができる。資料を公開したり原本画像を確認したりするには有用な電子化手法である。一方、デジタル画像であるため

計算機が電子化されたデータから漢文やそれに付与された訓点を認識することは難しい。そのため、計算機が一次資料化したデータを用いて統計処理を行うことは難しい [3].

二次資料化は計算機を用いた統計処理や自然言語処理といったデータ解析を行うのに適した電子化手法である。しかし、現在、WindowsOS や LinuxOS を搭載した一般的な計算機では漢文本文をテキストとして入力することは可能であるが、その漢文に付与された訓点を電子化する方式は定義されていない。そのため、訓点資料を二次資料化する場合、訓点を取り除いた漢文本文のみのテキストとして電子化するか、入力者が訓点を解釈し書き下し文としたものを電子化する方法がとられている。

漢文本文のみのテキストデータでは、訓点の情報が抜け落ちてしまい、訓点研究には利用できない。書き下し文のテキストデータでは、訓点情報が文章に変換されて保存されるため、一部の訓点研究には利用可能である。しかし、訓点の解釈は読者によって異なることが多い。原本画像から訓点を読み取り、書き下しを行った文章の場合、「入力者（読者）が読み取った文章」が入力されることになり、恣意が反映される余地が大きい。また、訓点は同じ位置・形状に付与されていても、時代や流派によって読み方が異なる。書き下された文章では、特定の時代・流派の読みに従って読み取られるため、異なる時代・流派の読みに従って読む場合は再度書き下し文を作成し電子化する必要がある。つまり、書き下しを行った文章は、「漢文本文」と「付与された訓点情報」に加えて「特定の時代・流派の読み」という情報が加えられた文章になってしまう。また、この書き下しや翻刻といった電子化に関する作業の煩雑さも問題となっている。これを支援するため田中ら [4], [5] は訓点資料の翻刻作業の支援を目的とした入力システムを実装している。このシステムは、実際の訓点資料の翻刻を目的としており、訓点資料のアーカイブ化を主目的とするものである。煩雑な入力作業を効率化できる点では非常に有効であり特定の漢文に付与された訓点を保存することは可能であるが、訓点そのものの解析に適した電子化方式ではないと考えられる。以上のように、現在の訓点資料の電子化手法では、訓点情報を残したまま計算機が処理可能な形式で電子化したり、訓点情報そのものを電子化したりすることができないという課題がある。

4. ヲコト点図の電子化

4.1 ヲコト点図電子データの構造

本研究では、訓点情報をテキスト化するための方式として、2.2 節で述べたヲコト点図を電子化するためのデータ構造を提案する [6]. 訓点情報は漢文のどの「位置」にどの「形状」の点が付与されているかを記録する必要がある。

4.1.1 ヲコト点の座標系

ヲコト点の多くは漢字の四隅と中心に打たれる。そこで

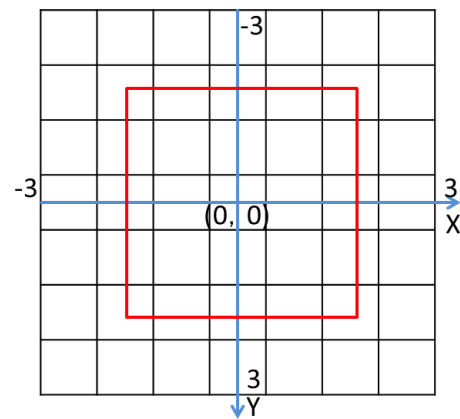


図 2 ヲコト点の座標系

Fig. 2 Coordinate system of Wokototenju.

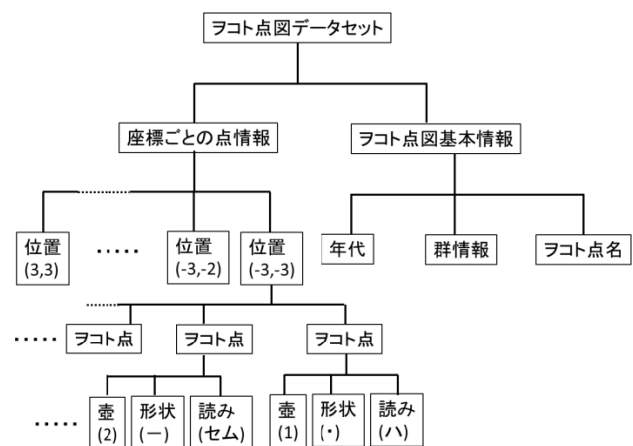


図 3 ヲコト点図データ構造概要

Fig. 3 Outline of a data structure of Wokototenju.

本研究では、四隅とそれぞれの辺の中心、さらにそこから 1 マス離れた座標を表現可能とする 7×7 マスのグリッド座標を用いる。また、中心 5×5 マスに示した四角形は漢字が書かれる範囲を示す。これにより、たとえば、「右上」にヲコト点が表示されている場合でも、漢字と重なっている位置なのか漢字から離れた位置なのかといった差異まで表現可能である。

今回設計した座標系を図 2 に示す。本座標系では x 軸 y 軸の 2 次元表記とし、字の中心を (0,0) 左上を (-3, -3)、右下を (3,3) とした。この座標系は、基点を漢字の中心に置くことで回転処理が行いやすくすることに主眼を置いている。ヲコト点は時代の変化によって点の位置（壺）が回転する特徴を持つという先行研究 [1], [2] にも適応しやすい形式である。

4.1.2 ヲコト点図の電子データ構造

ヲコト点図の電子化を行うために、出力する電子データは図 3 に示す木構造のデータ構造とした。東大寺点や喜多院点といったヲコト点の種類を 1 つの単位とし、データセットを作成する。1 つのデータセットには点情報と、そのヲコト点図の基本情報を持たせる。点情報は、7×7

表 1 代用した記号一覧 (形状は文献 [1] より引用)

Table 1 List of substituted symbols.

ヲコト点図中の形状	代用した組み合わせ
	十 十 (よこ)
	人 人 (よこ)
	二 七 (たて)
	二 母 (たて)
	┌ 十 (たて)
	┌ (たて)

マスのグリッド座標の位置情報を親とし、点の「形状」と「読み」、壺を子要素として持つ木構造のデータ構造とする。ヲコト点図の基本情報は、ヲコト点図の名前 (東大寺点, 喜多院点など), 使われていた年代, ヲコト点図の系統を表す群番号を定義し, 点図の種類を区別する際に用いる。

また, ヲコト点には 2.2 節で述べたとおり, 既存の文字集合内の文字では表現不可能な「形状」を持つ点も存在する。それらを表現するために特殊なフォントや画像データを使用した場合, テキスト解析時に処理が複雑化する。そのため, ヲコト点の「形状」は既存の文字集合内の文字で表現しテキストデータとして保存することとし, 表現不可能な「形状」については, 表 1 に示す「二母 (縦)」のように, 2 文字以上の組合せと, 組み合わせる方向 (縦, 横) をカッコ書きで表す方式で表現することとした。なお, 文字集合は Unicode を使用する。

この構造を保存するための電子ファイルは, 木構造を保持するのに適した XML 形式で記述する。XML で使用するタグの定義は表 2 のとおりとした。

4.2 ヲコト点図入力支援ツールの設計と実装

XML 形式のテキストデータは, 一般的にキーボードを使用して手入力を行うには煩雑な形式である。そこで, 提案するヲコト点図電子データ構造に従って, ヲコト点図の電子化作業を支援するツールを作成した [7]。本ツールは GUI を持ち, 直観的な操作によってヲコト点図を電子化する。

本ツールは, 主に以下の 4 つの画面から構成される。

1. メイン画面
2. 点図データ入力画面
3. XML 確認画面
4. 点情報確認画面

これらのうち, メイン画面の例を図 4 に, 点図データ入力画面の例を図 5 に示す。図 4 に表示された 49 個のボタンは, ヲコト点の座標軸 (7×7マス) に対応している。中心の 5×5 マスに表示されている赤い四角形は, 点図における漢字に見立てた壺を表す四角形, および漢字の中心

表 2 XML タグの定義

Table 2 Definition of the XML tags.

XML 要素名	内容
群番号	ヲコト点図の群番号を表す
ヲコト点図名	ヲコト点図の名前を表す
年代	ヲコト点が使われた年代を表す
data	ヲコト点図データの親要素である
grid	1 マスのグリッドを表す, data タグの子要素である
位置	グリッドの位置情報を表す, grid タグの子要素である
X	X 軸方向の位置情報を表す, 位置タグの子要素である
Y	Y 軸方向の位置情報を表す, 位置タグの子要素である
ヲコト点	一つのヲコト点を表す, grid タグの子要素である
形状	ヲコト点の形状を表す, ヲコト点タグの子要素である
読み	ヲコト点の読みを表す, ヲコト点タグの子要素である
壺	ヲコト点の壺を表す, ヲコト点タグの子要素である

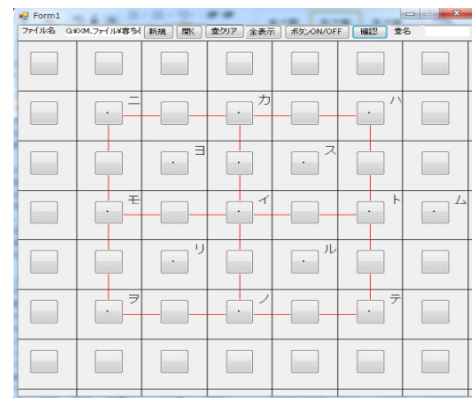


図 4 支援ツールメイン画面

Fig. 4 Main screen of the support software.

を視覚的に認識しやすくするための補助線である。ヲコト点の情報を入力したい座標にあるボタンをクリックすると図 5 に示す点データ入力画面が表示される。点データを入力すると, 図 4 に示したメイン画面のボタンに, 入力した点の「形状」が表示される。また入力した点の「読み」はセルの右上に表示される。本ツールでは, ヲコト点図は壺ごとに入力すると想定している。図 4 のメイン画面における入力データの表記は 1 つの壺を入力し終わった状態である。その表現方法は, 図 1 中に示した壺の 1 つと同じような表現になるようにメイン画面の設計・実装を行った。

表 4 ヲコト点の登場回数と点の種類

Table 4 Appearance number of Wokototen and the kinds of Wokototen.

登場回数	点の種類	登場回数	点の種類
37	4	18	4
36	0	17	5
35	2	16	4
34	3	15	2
33	1	14	6
32	2	13	7
31	2	12	3
30	2	11	10
29	4	10	9
28	1	9	6
27	5	8	9
26	4	7	11
25	1	6	16
24	2	5	10
23	1	4	18
22	4	3	33
21	2	2	84
20	3	1	310
19	4		

ト点の読みが何種類あったかをまとめた結果を表 4 に示す。結果から、読みの種類 594 種のうち、半数以上の 310 種（すべての読みのうち 52.2%）が登場回数 1 回、つまり 1 つのヲコト点にしか存在しない読みであることが分かった。登場回数が 2 回、3 回の点も数が多く、それぞれ 84 種、33 種であった。登場回数が 1~5 回であるヲコト点の読みは合計で 455 種であり、ヲコト点の読みの種類の 76.6% を占めることが分かった。

次に、ヲコト点の「読み」の種類ごとに出現回数を計測した。594 種の読みの中で主要ヲコト点 26 種に最も多く出現したものは「ス」「ナル」「ナリ」「タリ」であり、出現回数はそれぞれ 37 回であった。次に多く出現したものは「ヨリ」「シ」であり 35 回であった。解析結果のうち、登場回数が 20 回を超えたヲコト点の読みを表 5 に示す。ここで、入力したヲコト点図の総数である 26 種よりも登場回数の多いものが存在するが、それは 1 つのヲコト点図の中で「同一の読みを表すが形状が異なる点」が存在するためである。たとえば、仁都波迦点には「ス」と読むヲコト点「・」と「一」の形状で 2 つ存在する。

図 7 はヲコト点の読みの長さの平均を登場回数ごとに求め、表 4 の結果とあわせてグラフ化したものである。読みの登場回数別に測定した読みの平均文字長を見てみると、登場回数 1 回の読みが 4.1 文字と最も長いことが分かる。

表 5 登場回数 20 回以上のヲコト点の読み

Table 5 “Yomi” of Wokototen (More than 20 times appeared).

読み	登場回数	読み	登場回数
ナリ	37	カ	27
ナル	37	ハ	27
ス	37	セム	27
タリ	37	テ	27
ヨリ	35	ニ	26
シ	35	リ	26
アリ	34	タマフ	26
タル	34	ム	26
セリ	34	ケリ	25
ト	33	ヨ	24
コト	32	タテマツル	24
セ	32	レ	23
モ	31	アル	22
ヲ	31	シテ	22
キ	30	ラ	22
ク	30	ナ	22
ノ	29	ヤ	21
スル	29	フ	21
ヌ	29	モノ	20
セル	29	ケ	20
ル	28	ツ	20
ソ	27		

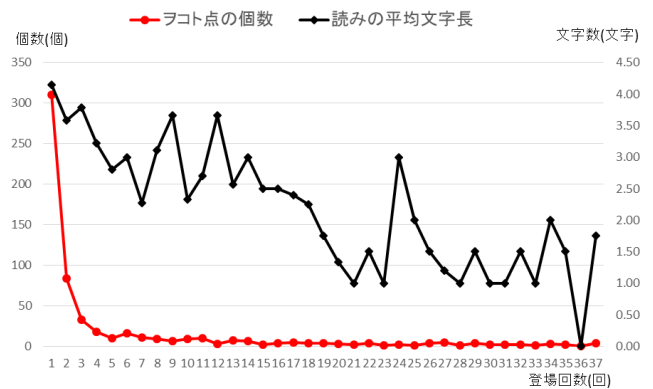


図 7 ヲコト点の登場回数と読みの平均文字長
Fig. 7 Appearance number of Wokototen and the length of the word.

表 6 に登場回数が 1 回のヲコト点の読みの例を、表 7 に登場回数が 2 回のヲコト点の読みの例を示す。これらから、登場回数が多い表 5 に示された読みと比較して、登場回数の少ない読みは文字長が長い。登場回数が 1 から 18 回までの読みは、平均文字長が 2 文字を超えており、登場回数 19 から 37 回までの読みは、登場回数 24 回を除くと平均文字長 2 文字以下となっている。登場回数 24 回の文字は表 5 に示すとおり「ヨ」と「タテマツル」の 2 種である。

表 6 登場回数 1 回のヲコト点の読みの例
Table 6 Example of “yomi” (appeared once).

ヲコト点の読み		
ヲカニ	ナラムル	ハムヘリ
クル	セイマス	トニハ
トイハムカコトク	サタメテ	マム
トイハムトニアラシ	コトヽク	ヘク
シハ	ニオイテ	サヘニ
サマニ	シカレトモ	トイフヨリ
サロ	トイフトソ	タルカ
イフナリ	セシメシ	オモヒキ
トノコトク	タデマツルニ	

表 7 登場回数 2 回のヲコト点の読みの例
Table 7 Example of “yomi” (appeared twice).

ヲコト点の読み		
トイフナリ	セラハ	タリキ
ハヘリ	ハム	タラハ
ヲハ	イハ	フルコト
シメキ	サヘ	コトモ
イフコト	ルコト	マル
シク	テム	スルトキ
トナリ	イヘル	オモヘリ
ミモトニ	ヘルイ	タマヘキ

表 8 位置別のヲコト点登場回数
Table 8 The appearance number of Wokototen (according to the location).

X \ Y	-3	-2	-1	0	1	2	3
-3	0	0	0	0	0	0	0
-2	1	332	7	284	7	337	1
-1	0	9	32	5	28	14	0
0	1	279	12	242	13	292	13
1	0	5	33	2	26	9	0
2	0	327	12	275	14	331	1
3	2	0	0	0	0	0	0

5.1.3 位置に関する基礎計量

ヲコト点は、同一の形状を持つ場合でも点が付与される「位置」によって読みが異なる。ヲコト点は漢字の四隅に付与されることが多いが、実際にどの位置に付与される点が多いか、ヲコト点の読みの種類ごとに主要ヲコト点 26 種のデータを用いて求めた。結果を表 8 に示す。表中の赤く示した部分は、ヲコト点図において漢字があると想定さ

位置別のヲコト点登場回数

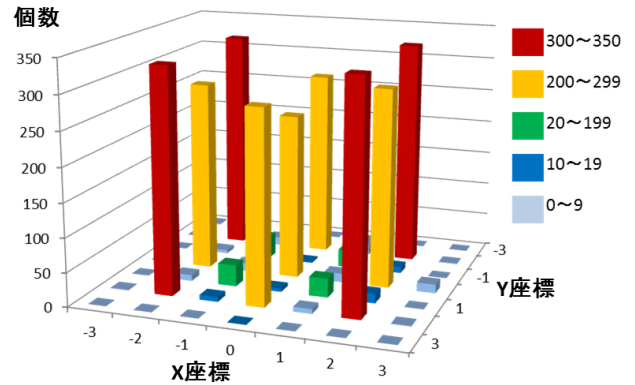


図 8 位置別のヲコト点登場回数

Fig. 8 Graph of the appearance number of Wokototen (according to the location).

れる壺の枠に対応する部分である。また、表 8 の値をグラフ化したものを図 8 に示す。

これらの結果から、漢字の四隅に該当する $(-2, -2)$, $(2, -2)$, $(-2, 2)$, $(2, 2)$ と、漢字の上辺下辺左辺右辺のそれぞれ中央 $(-2, 0)$, $(0, -2)$, $(0, 2)$, $(2, 0)$, および漢字の中心 $(0, 0)$ の合計 9 カ所にヲコト点が多く付与されることが分かる。これら 9 カ所の中で最も多くヲコト点が付与された位置は、漢字の四隅の右上 $(2, -2)$ であり、その数は 337 個であった。次に多いのは漢字の四隅の左上 $(-2, -2)$ の 332 個、右下 $(2, 2)$ の 331 個、左下 $(-2, 2)$ の 327 個であった。一方、漢字の内部と漢字の外部にもヲコト点が付与されることが分かる。そのうち、中心 $(0, 0)$ を除く漢字の内部に打たれる点と、漢字の外部に打たれる点では、内部に打たれる点の方が、種類は多いという結果になった。その中でも、左下側の $(-1, 1)$ の位置が 33 種類と最も多い。漢字の内部に限ると、中心 $(0, 0)$ の上下にあたる座標 $(0, 1)$, $(0, -1)$ は付与されるヲコト点の種類が少ないことも分かる。また、漢字の外部に限ると、漢字の右側の中心座標 $(3, 0)$ に最も多くの種類のヲコト点が付与されることが分かる。さらに、漢字の上下では $(-3, 3)$ に 2 種類の点が付与されていた以外は、ヲコト点がまったく付与されていないことも分かる。

5.1.4 形状に関する基礎計量

代表的なヲコト点の「形状」は星点「・」だが、それ以外にも様々な形状が存在する。ここでは、主要ヲコト点 26 種中に、何種類の形状が存在するかを調べた。

解析の結果、67 種類の形状が存在することが分かった。登場回数が多かった形状の上位 20 種を表 9 に示す。最も多く登場した形状は「・」であり、その回数は 337 回であった。2 番目に多い形状は「┌」で 250 回であった。形状「・」は「┌」よりも 87 回多く登場しており、ヲコト点図においては形状「・」が最もよく使われることが定量的に示された。また、登場回数が 200 回を超える形状は「・」

表 9 登場回数上位 20 種 (形状別)

Table 9 List of “Keijo” in order of the amount appearance.

形状	登場回数	形状	登場回数
・	337	+	67
ㄥ	250	∴	65
	242	└	64
—	242	=	62
＼	242	リ	50
ㄣ	220	ソ	46
/	211	┘	39
人	125	⊥	35
フ	109	イ	32
:	98	∨	28

表 10 共通して使用されるヲコト点数上位 20 種 (形状別)

Table 10 List of the number of the common “Keijo”.

形状	共通するヲコト点数	形状	共通するヲコト点数
・	26	∴	10
ㄥ	26	∨	10
	26	:	9
—	26	=	9
＼	25	リ	9
ㄣ	25	└	8
/	23	┘	8
フ	19	⊥	8
人	15	ナ	7
+	11	○	6

「ㄥ」「|」「—」「＼」「ㄣ」「/」の 7 種類であった。

次に、67 種類の形状が、何種類のヲコト点で使われているかを調べた。主要ヲコト点 26 種それぞれのヲコト点に 1 度でも登場すれば 1 つと数え、26 種のヲコト点のうちいくつに共通して使用されているか (最大 26) を調べた。使用されているヲコト点が多かった形状の上位 20 種類の結果を表 10 に示す。すべてのヲコト点に共通して使用されていたものは「・」「ㄥ」「|」「—」であった。25 種類のヲコト点に共通して使用されていた「＼」と「ㄣ」はそれぞれ、智証大師点と池上阿闍梨点に存在しなかった。

5.2 考察

5.2.1 読みに関する考察

5.1.2 項より、多くのヲコト点図に登場するヲコト点は、助詞、助動詞が多いことが分かる。登場回数が多いヲコト点は、漢文を訓み下すときに高い確率で必要となるヲコト点であると考えられる。よって、実際の訓点資料にも多く記載があるものと推測され、これらは重要なヲコト点であ

ると考えられる。また、よく使われる点は漢文訓読において多用する点であり、多くの点図に記載されていると考えられる。これらの点は、当時の漢文読者・学習者がある流派のヲコト点を使用するためには、必ず学習することになる点であり、どういった形状でどこ位置に付与されるかは、そのヲコト点を特徴づける重要な情報であると考えられる。

次に、表 6、表 7 および図 7 より、登場回数の少ない読みには、複合語が多くみられることが分かる。登場回数が 1 から 18 回までの読みは、平均文字長が 2 文字を超えていることから、複合語が中心であると考えられる。登場回数が 19 回以上の読みは、24 回を除くと平均文字長が 2 文字以下であり 1 音節、2 音節の読みが中心だと考えられる。たとえば「トイフナリ」や「トイハムカコトク」などが該当する。前者は「ト」と「イフ」と「ナリ」を組み合わせた読みである。「トイハム」は、主要ヲコト点 26 種に 11 回登場し、また「ナリ」は 37 回登場しているが、「トイフナリ」は 2 回しか登場しない。登場回数 24 回の読みの平均文字長は 3 文字であり比較的長いという結果が得られている。これは、該当する読みが「ヨ」と「タテマツル」の 2 種のみであり、「タテマツル」が 5 文字であり、その影響が大きく平均文字長が長くなっているためである。このことから、登場回数が多い読みはその種類が少ないため、平均文字長を計算したときに外れ値が生じやすいものと考えられる。

5.2.2 位置に関する考察

ヲコト点は、漢字の四隅に付与されることが多いといわれているが、5.1.3 項の結果からそれが事実であることが定量的に示された。中でも漢字の右上 (2, -2) と右下 (2, 2) に付与されるヲコト点の種類が最も多いことから、ヲコト点図が作られるときには、漢字の右側から読みと形状が決められていくのではないかと考えられる。この計量結果から得られる推測は、築島 [9] の“最近の石塚晴通氏の研究によれば、唐土では、声点を漢字の字面に附記するのに、右上隅から平声を始めた例もあつたやうであつて、声点とヲコト点との間に関連があると考へるならば、右上をテとして、それを基点とする形が寧ろ古いのではないかとも思はれる”とするヲコト点基点論とも一致する。このような人文科学分野における研究課題に対し、今回の基礎計量の結果を用いて推測を行うことは今後の重要な課題となると考えられる。

漢字の外部に付与される点の計量結果から、ヲコト点は漢字の右側には比較的付与されやすいが、上下にはあまり付与されていないことが分かる。これは、漢字の上下に点を付与した場合、続いている 2 漢字のどちらに付与してある点が判断しにくいためではないかと考えられる。

5.2.3 形状に関する考察

5.1.4 項の結果から、ヲコト点は代表的な形状である星点

「・」が、最も種類が多いことが分かる。このことから複数のヲコト点図の関係性を比較する場合には、多くのヲコト点図に登場する星点「・」を中心に比較を行うのが効率的かつ有効であるといえる。さらに、「・」「ㄣ」「|」「—」の4種類は主要ヲコト点26種すべてに存在するため、これらの形状を中心にヲコト点図の比較を行うことで、同一の形状が存在しないため比較できないという事態にはおちいらぬ。また、2つのヲコト点どうしのように少数のヲコト点を比較する場合には、登場回数の少ない形状を持つ点が共通しているかどうか、といった観点から比較することも有用であると考えられる。

主要ヲコト点26種すべてに存在する「・」「ㄣ」「|」「—」は、すべて一筆書きで付与できる形状である。それ以外にも「\」「ㄣ」「/」など、20種以上の主要ヲコト点に共通して使用されている形状は、一筆書きで付与できる形状である。このことから、簡単に書ける形状ほど、多くのヲコト点に使われていることが分かる。

5.2.4 提案構造化記述方式の有効性

5.1.1 から 5.1.4 項では、本研究で提案したヲコト点の構造化記述方式を用いて電子化した主要ヲコト点26種の電子化テキストを利用している。ヲコト点図に記載されたヲコト点を、7×7マスの提案座標系にすべて配置し、電子化および基礎的な計量を行えたことから、本方式は訓点の電子化にとって十分に有効な構造化記述方式であると考えられる。

また、本研究では実際の訓点資料は対象とせず、主要ヲコト点26種に記載されたヲコト点を電子化している。そのため、特定の漢字に依存せず、訓点のみの情報を保存可能な方式であるといえる。このように訓点データと漢文本文とを分離して記述できる方式となっているため、訓点に関する計量が容易である。XML形式のテキストデータとして保存されているため、本方式で電子化したデータを用いれば点図データベースなどの作成も容易であると考えられる。

本方式で問題となる点は、ヲコト点の形状の表現についてである。ヲコト点の形状はUnicodeの文字集合内の文字から似通った形を持つ文字を選んで代用している。そのため、実際の点図に記載されている形状とは若干の違いがある。また表2に示すように、一部代用できない形状が存在し、複数の文字を組み合わせることで表現する方式をとっているため、それらの文字の実際の表記との違いが大きくなっている。なかでも卍を「十」(よこ)と横向きの組合せを表現する場合は視覚的にも比較的分かりやすいが、毛を「二七(たて)」のように縦向きの組合せを表現する場合、視覚的に分かりにくく実際の点図との表記差も大きくなる。計算機上での表記と実際の資料における表記との差は、字形の研究を行ううえでは問題となりうる。そのため、計算機を用いた訓点研究を進めていくうえでは、訓点記号

を表す文字コードの定義、提案も課題の1つとなると考えられる。

6. まとめ

本研究では、訓点の一種であるヲコト点を、計算機を用いて解析することを目的としヲコト点の構造化記述方式を提案した。提案方式に基づいて、ヲコト点図の電子化を支援する入力支援ツールを開発し、主要ヲコト点26種を電子化した。さらにそれを用いてヲコト点の基礎計量を行った。その結果、ヲコト点の総数、主要ヲコト点26種の中でヲコト点の読みは「ス」「ナル」「ナリ」「タリ」が最も多く登場し、ついで「ヨリ」と「シ」が多いこと、ヲコト点は漢字の四隅とその間および中心に多く付与されること、最も多く登場する形状は星点「・」であることを定量的に示した。また、実際に26種のヲコト点を電子化・計量することで、7×7マスの提案座標系にすべてのヲコト点を配置可能であり、電子化データをヲコト点の解析に利用可能であることを確認し、本提案方式の有効性を確認した。

今後の課題として、提案方式を用いて電子化したデータをヲコト点どうしの関係性の解析や、ヲコト点データベースの作成などの訓点研究に役立てることが必要であると考えている。また、本提案方式をもとに、漢文データと組み合わせる表記可能な構造化記述方式を検討し、現存する訓点資料をテキストベースで訓点情報を損なうことなく電子化する。さらに、その情報を用いて訓点コーパスを作成することも今後の課題の1つと考えている。

謝辞 本研究はJSPS科研費15H06833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] 築島 裕：訓点語彙集成<第1巻>、ヲコト点概要、汲古書院(2007)。
- [2] 中田祝夫：古点本の国語学的研究、講談社(1954)。
- [3] 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)。
- [4] 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および評価、第15回情報科学技術フォーラム(FIT2016)、pp.7–14 (2016)。
- [5] 田中 勝、村川猛彦、宇都宮啓吾：訓点資料における翻刻支援システムの構築、人文科学とコンピュータシンポジウム(じんもんこん2015)、Vol.2015, pp.263–268 (2015)。
- [6] 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」、じんもんこん2012論文集2012, Vol.2012, pp.211–216 (2012)。
- [7] 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」、じんもんこん2015論文集, Vol.2015, pp.185–190 (2015)。
- [8] 堤 智昭、田島孝治、高田智和、小助川貞次：コンピュー

タを用いた主要ヲコト点の関係性の解析, 人文科学とコンピュータシンポジウム「じんもんこん 2016」, じんもんこん 2016 論文集, Vol.2016, pp.139-146 (2016).

[9] 築島 裕: 平安時代語新論, 東京大学出版会 (1969).



堤 智昭 (正会員)

東京電機大学. 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情報工学専攻博士前期課程修了. 2015 年同大学院工学府電子情報工学専攻博士後期課程修了. 現在, 東京電機大学情報環境学部助教.

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).



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).



小助川 貞次

富山大学. 1981 年北海道大学文学部卒業. 1986 年同大学大学院文学研究科修士課程修了. 1987 年同大学文学部助手の後, 1993 年より富山大学人文学部助教授を経て, 現在, 同大学同学部教授. 自言語による古典語文献読解

に関する比較研究, 日本国内現存漢籍訓点資料の解読・記述研究に従事. 日本語学会, 訓点語学会各会員. 文学修士.



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).