

訓点データベースを用いたヲコト点図の機械的分類手法の検討

堤 智昭 (筑波大学)
田島 孝治 (岐阜工業高等専門学校)
高田 智和 (国立国語研究所)
小助川 貞次 (富山大学)

概要: 本稿では訓点データベースを活用し、訓点資料に記載されているヲコト点を一覧にしたヲコト点図を機械的に分類する手法について検討を行う。築島裕の平安時代訓点本論考に記載された点図のうち、第1群から第8群までに分類されている159種を対象に、それぞれの点図がどの群に分類されるものであるかの機械的な分類を試みた。その結果築島裕の既存研究で分類された群を正解とした場合、159点図中153点図が正解の群に分類できた。また、特殊点と呼ばれる、第1群から第8群に分類されない点図についても分類を行うことができた。一方で、四隅や字画の中心といった重要とされる位置以外の点を用いた場合、従来研究の群分類とは25%程度の一致に留まることがわかった。

キーワード: 訓点資料, ヲコト点, 自動化

A Study on a Mechanical Classification Method of Wokototenzu Using the Kunten Database

TOMOAKI TSUTSUMI (University of Tsukuba)
KOJI TAJIMA (National Institute of Technology, Gifu College)
TEIJI KOSUKEGAWA (University of Toyama)
TOMOKAZU TAKADA (National Institute for Japanese Language and Linguistics)

Abstract: In this paper, we examine a method for mechanically classifying the wokototenzu using the data in kunten database. We have attempted to mechanically classify the 159 Wokototenzu in previous studies by Hiroshi Tsukishima, which are classified groups 1 to 8. As a result, 153 out of 159 Wokototenzu were classified into the correct group if the group classified in the existing study by Hiroshi Tsukishima was considered as the correct answer. In addition, we were able to classify special points, which are not classified groups 1 to 8. On the other hand, when we used Wokototen other than those considered important, such as the four corners or the center of a character, the classification did not agree well with the group classification of previous studies.

Keywords: Kunten Material, Wokototen, Automation

1. はじめに

かつて東アジア諸国で中国語である漢文が広く受容されていた。漢文が記載された資料を、当時の人々が自国の言語として内容を理解するために、符号や文字を用いた訓点を漢文本文に直接加える読書方法が行われてきた。このときに用いられたものが訓点であり、訓点の書き入れがなされた資料が訓点資料である。訓点資料は主に日本語の表記、音韻、語彙、文法などの日本語史の研究に利用されている。我々はこれまでに、これら訓点を電子的に取り扱うための方法について提案を行い、訓点データベースとしてまとめた。本稿では、この訓点データベースを活用し、訓点資料の自動書き下しや、ヲコト点図の自動的な種類判別を目的とし、訓点資料から作成されたヲコト点図を機械的に分類する手法について検討を行う。

2. ヲコト点図とその分類

ヲコト点は漢字の字画内部や周辺につけられる符号であり、その形状（・ー┐└等）と付けられた位置によって読みが決まる。どの位置にどのような形状の点を付与するかは、ヲコト点を使用する流派や加える者によって様々な種類が存在する。

このヲコト点を、形状ごとに「壺」と呼ばれる漢字に見立てた四角形の周辺に、その読みとともに記載しまとめたものをヲコト点図という。代表的なヲコト点を種類ごとにまとめたヲコト点図は、訓点を読み解くための資料として編纂され、歴史的に伝承されている。また、訓点資料を解読し日本語文に書き下す過程において、その訓点資料に記載されているヲコト点を一覧にしたヲコト点図が作成されることもある。

このように一言にヲコト点図といっても二種

類のものが存在する。訓点資料を読み解くための解釈の基準となる点図と、実際の訓点資料を研究し読み解くための、読者の解釈に基づく点図の二種類である。本来これらを個別に区別するような名称は訓点研究の中では存在しないが、本稿では便宜上、伝承されてきたヲコト点の解釈を示すヲコト点図を「ヲコト点図（スキーマ）」、実際の訓点資料を解釈し作成されたヲコト点図を「ヲコト点図（解釈）」と称する。

ヲコト点図（スキーマ）には、喜多院点や西墓点といった代表的なヲコト点図 26 種（以下、主要ヲコト点図 26 種）が存在する。これらは中田祝夫[1]や築島裕[2]によってまとめられ、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）に分類されている。これらの分類は、歴史の変遷を考慮し、ヲコト点の中でも用いられることが多い、字画の四隅に付与された星点「・」を基準に分類方法が検討されている。例えば、図 1、2 に示す第 5 群点に分類されている円堂点と乙点図の星点「・」を比較してみると、四隅が左下から時計回りにテ・ニ・ヲ・ハと同じ読みが記載されていることがわかる。

一方、ヲコト点図（解釈）には、東大寺図書館が所蔵する法華文句を解釈し、そこに記載されていたヲコト点をまとめた点図、というように読者の解釈に基づいていくつも点図が作られている。これらはその訓点資料の特性を表す重要な情報であり研究成果であると言える。実際に、訓点資料を対象とした主な研究[1][3][4][5][6][7]では、訓点資料研究の成果を発表する方式として、漢文史料に付与されたヲコト点や仮名点を解読して日本語文にする書き下し文方式が多く取られており、その過程において作成されたヲコト点図も研究成果とされる場合が多い。

また、訓点研究においては、訓点資料に基づく点図を分析するために、訓点の内容がどの流派に近いものかを判別することが行われてきた。その時に基準として用いられるのが、ヲコト点図（スキーマ）である主要ヲコト点図 26 種や、それらを第 1 群から第 8 群までに分類した群分類である。

3. 分類するヲコト点図

今回分類を試みる対象とする点図は、築島裕[2]記載の点図 191 種とした。築島裕[2]記載の点図 191 種のうち 159 種はどの群に分類される点図であるか判定されている。残りの 32 種は特殊点と呼ばれる点図である。特殊点は、平安時代初期ごろの古い訓点資料に使用されたため、第 1 群から第 8 群には分類できないとされている。

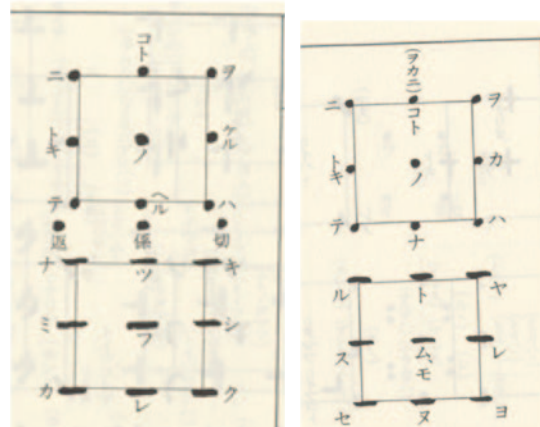


図 1 円堂点の一部 ([2]より引用) 図 2 乙点図の一部 ([2]より引用)

実験には、これらの点図を電子化したヲコト点データベースのデータを用いる。データベースに入っているデータ項目は、ヲコト点図（スキーマ）とヲコト点図（解釈）では若干の違いがあるが、今回は両方に共通している、ヲコト点の「形状」と「読み」、つけられた位置の「X 座標値」「Y 座標値」、その点図の「種類」は何か、の項目を利用する。「形状」「読み」「種類」のデータは文字形式で保存されている。ヲコト点図がつけられた位置は原点を (0,0) とする 7×7 マスの座標で表現されている。具体的には「X 座標値」「Y 座標値」それぞれに、-3 から 3 までの整数値でデータベースに保存されている。

4. ヲコト点図の分類方式

本研究では、ヲコト点図を機械的に分類するために適切な手法を検討することを目指して実験を行う。そこで、まずは従来の研究成果を正解データとし、ヲコト点図データベースに記載されているヲコト点図（解釈）のうち、1 群から 8 群までに分類されている点図の分類を行う。これらの点図を分類した時の正解は、既存研究で分類されたものを正解として、どの程度の精度で機械的に分類できるかを求める。

ヲコト点図は図 3 に示すように、どのような読みのヲコト点か漢字のどの部分に、どのような形状でつくかを記載したものである。文章に関する解釈を表現するための情報であるため、その解釈に関する内容を分類しようとした場合、図として処理し画像内から特定のパターンを検出するパターンマッチングのような手法で類似性を求める処理を行うことは不適切であると考えられる。しかし一方で、漢字に見立てた四角形を用いてどの位置に点がつけられているのかという情報も意味を持つため、単純に文字列を比較する場合と同じように考えることもできない。そこで本研究では、2 種類の方法で分類対象の点図の分類を試みた。



図3 喜多院点の点図例 ([2]より引用)

(1) 一致したフコト点の数に基づく判定方法

1つ目の方法は、同一の「位置」「形状」「読み」を持つ点が多いほど類似性が高いとする計算方法による分類である。ヲコト点図（解釈）のデータ一つ一つと、ヲコト点図（スキーマ）のデータである主要ヲコト点図 26 種とで比較、分類する。計算の結果、最も類似しているとされた主要ヲコト点図が属する群が、その点図の判定結果となる。例えば、ある資料から作られた点図が主要ヲコト点図 26 種のうち第 2 群点である喜多院点に対して最も多く一致するヲコト点を持っていた場合、その点図は第 2 群点であると分類される。

ヲコト点図を分類するための適切な判定方式を検討するため、単純に点図に記載された全てのヲコト点を比較するだけでなく、比較する「形状」や「位置」を変化させて複数の計算方法で実験を行った。「形状」については、ヲコト点図の中でよく使われる「・」「┐」「|」「—」の4つの形状をそれぞれ用いて比較した場合と、形状を指定せずに分類した場合で精度を比較した。これら4つの形状は、主要ヲコト点図26種には全ての点図に記載がある形状である。形状による比較条件をまとめると以下の通りとなる。

- i. 全ての形状の点を対象とする
- ii. 「・」のみを対象とする
- iii. 「┐」のみを対象とする
- iv. 「|」のみを対象とする
- v. 「—」のみを対象とする
- vi. 「・」「┐」「|」「—」を対象とする

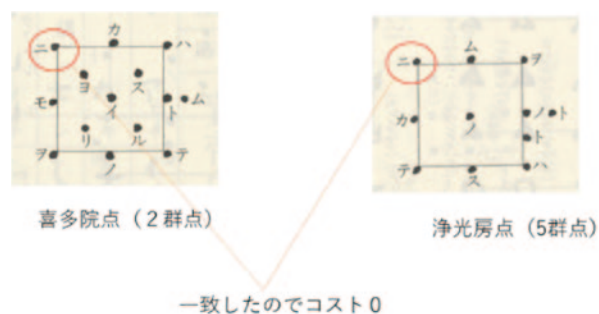


図 4 編集距離コスト (一致の場合)

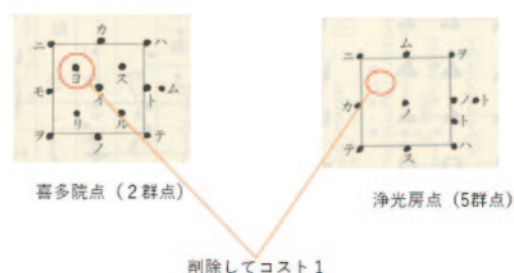


図 5 編集距離コスト（追加・削除の場合）

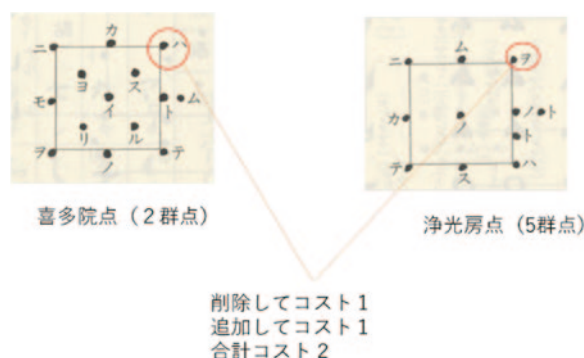


図 6 編集距離コスト (置き換えの場合)

「位置」については、ヲコト点がよく付与されるとされる、字画の四隅、上辺中央、下辺中央、左辺中央、右辺中央、および中央にある点の9つを重み付けした場合と、重みをつけない場合とで精度の比較を行った。この時、形状は星点「・」に限る(条件ii)こととした。「位置」による比較条件は以下の通りである。

- ①全ての座標を平等に1倍とする
②9箇所のうち、四隅を3倍、中心を2倍、9箇所以外を1倍とする
③9箇所のみ1倍、それ以外は0倍とする
④9箇所のみ0倍、それ以外は1倍とする

(2) ヲコト点図の編集距離に基づく計算方法
次に、文字列の比較を行う編集距離の考え方を元に、ヲコト点図同士の編集距離を計算し、距離が近い主要ヲコト点図が属する群を、そのヲコト点図の群とする方法による分類を行った。一般的

な文字列における編集距離計算では、比較元の文字列が比較対象の文字列に書き換わるまでに必要な編集工程をコストとして計算し、それを編集距離とする。編集距離が短いほど、2つの文字列が類似していると考えられる。

今回は単純化のために形状は星点「・」のみを対象とすることとした。そのうえで、ヲコト点図の編集距離として、49マスそれぞれの点を比較し、追加、削除をそれぞれ1コストとして計算することとした。例えば、第2群点の喜多院点と第5群点の浄光房点を比較した場合、図4、5、6にあるような、以下の3パターンが考えられる。

1. 図4のように同じ位置に同じ読みの点があった場合はコスト0とする。
2. 図5のように片一方に点がなく、もう片一方にある場合は追加・削除のどちらかを行うためコスト1とする。
3. 図6のように両方に点がありかつ読みが異なる場合は、削除と追加を両方行うためコスト2とする。

5. 分類結果

(1) 一致したヲコト点の数に基づく判定方法による分類結果

まず、既存研究により第1～第8群に分類されている築島裕[2]記載の点図159種に対して、一致したヲコト点の数に基づく判定方法により分類を試みた。その結果を表1、表2に示す。

形状による違いは、i、ii、viで90%を超える正答率を得た。特に、ii、viは95%に近い正答率となっている。一方で「・」以外の形状を用いたiii、iv、vの場合では20%程度の低い正答率となり、判別できない点図も多く存在した。

位置による違いは、①②③で95%を超える高い正答率を得、いずれも近い結果となった。①と②③で結果が異なった点は1つであり、聖語蔵の四未曾有経であった。この点図は築島の分類では第4群点に分類されるものであるが、①の判定方法では第2群点となり、②の判定方法では第4群点となった。また、9箇所にあるヲコト点を用いない④では25%程度の低い正答率となった。

次に、築島裕[2]記載の特殊点32種を第1群から第8群に分類した。先に行った159種の結果を受けて、使用する形状は「・」点のみ(条件ii)と、全ての形状を対象とする(条件i)の2方式とした。

条件iのとき、32種の点図のうち、位置に関する条件が①②③の場合は31種の点図がいずれかの群に分類できた。主要ヲコト点図26種のいずれとも一致する点の一つも存在せず、分類できなかった点図は聖語蔵の「大方廣佛華嚴經十七卷」の1つである。条件④の場合は23個の点図において一致する点の一つも存在せず分類できなかった。条件i-①のときに第1群から第8群まで、

表1 分類結果1 (形状の違い)

判定方法	築島分類と合致した点図数 (正解判定数/総数)	判定候補が複数群であった点図数
i	144/159	5
ii	152/159	6
iii	28/159	27
iv	33/159	52
v	25/159	33
vi	151/159	11

表2 分類結果2 (位置による違い)

判定方法	築島分類と合致した点図数 (正解判定数/総数)	判定候補が複数群であった点図数
①	152/159	6
②	153/159	4
③	153/159	4
④	41/159	32

表3 特殊点の分類結果 (条件i-①)

分類された群	点図数
第1群	9
第2群	6
第3群	2
第4群	6
第5群	6
第6群	0
第7群	0
第8群	2

それぞれいくつの点図が分類されたかを表3に示す。また、具体的な分類結果のうち一部を表4に示す。

条件iiのとき、32種の点図のうち、位置に関する条件が①②③の場合は主要ヲコト点図26種のいずれとも、一致する点の一つも存在せず分類できなかった点図は5つである。これら5つの点図について、表5に示す。残りの27種の点図は、最低一つ以上は一致する点があり、いずれかの群に分類ができた。条件④の場合は23個の点図において一致する点の一つも存在せず分類できなかった。

(2) ヲコト点図の編集距離に基づく計算方法による分類結果

(1)と同様に、まず、既存研究により第1～第8群に分類されている築島裕[2]記載の点図159種に対して、編集距離に基づく判定方法により分類を試みた。その結果、159点図中142点図が正解データとした築島の分類と一致した。それらの中で編集距離が最小のものは2、最大のものは22、標準偏差は4.0であった。一方、一致しなかった17点図の編集距離は、最小のものが5、最大のものが16、標準偏差は3.5となった。また、特殊点32種は最小のものが7、最大のものが21、標準偏差は3.6となり、一致したヲコト点の数に基づく判定方法の条件i-①のときの判定結果と一致

表 4 特殊点の群分類結果 (条件 i -①, 一部抜粋)

訓点資料名	所蔵	書写年代	加点年代	類似判定された 26 種ヲコト点	群分類	一致 数
十八契印法并儀軌次第	石山寺	平安時代中期	平安時代中期	西墓点	第 1 群	7
菩提心論	石山寺	長保 [999-1004 年]頃	長保 [999-1004 年]頃	甲点図	第 1 群	6
大毗盧遮那成佛經〔第一次点〕	国立国会図書館	奈良時代	平安時代中期	喜多院点	第 2 群点	6
宝楼閣經	石山寺	平安時代初期	平安時代初期	仁都波迦点	第 1 群点	6
彌勒經疏一卷	酒井宇吉氏	唐時代	平安時代初期	順暁和尚点	第 8 群点	5
彌勒經疏三卷	園城寺	唐時代	寛平 2 年	甲点図	第 1 群点	5
瑜伽師地論〔第四次点〕	石山寺	天平 2 [730] 年	平安時代初期	西墓点	第 1 群点	5
蘇悉地羯羅供養法	大東急記念文庫, 高山寺	康平 7[1064] 年	康平 7[1064] 年	甲点図	第 1 群点	5
大智度論〔第五次点〕	石山寺	奈良時代	平安時代初期	広隆寺点	第 4 群点	5
華嚴經探玄記〔第一次点〕	聖語蔵	奈良時代	平安時代初期	天爾波留点別流	第 4 群点	4
瑜伽師地論〔第三次点〕	石山寺	奈良時代	平安時代初期	妙法院点	第 1 群点	4
大方廣佛華嚴經〔第二次点〕	石山寺	奈良時代	平安時代初期	喜多院点	第 2 群点	2
成唯識論述記〔第二次点〕	知恩院	平安時代初期	平安時代中期	古紀伝点	第 5 群点	2
羅摩伽經	聖語蔵	768 年	平安時代極初期	順暁和尚点	第 8 群点	1
大方廣佛華嚴經十九卷	聖語蔵	奈良時代	平安時代極初期	一致なし	分類不能	0

したものは 12 点図, 一致しなかったものは 20 点図となった。

6. 考察

5 章の (1) の結果から, 一致したヲコト点の数に基づいて主要ヲコト 26 種と比較を行った場合, 形状「・」の点 (星点) のみを対象とし, 四隅や字面の中心といった重要とされる 9 箇所を点を対象とすれば, 従来研究である築島の群分類と高い確率で分類可能であると言える。このことから, 点図を作成する元となった訓点資料に関する情報 (書写年代や加点年代等) を用いずとも, 従来の群分類と同様の結果を機械的に得ることが可能であると考えられる。

一方で, 星点以外の形状や, 四隅や字面の中心など重要と言われている位置以外についてヲコト点のみを用いて分類を行った場合, 従来の群分

類と一致する結果が得られづらい事がわかる。そのため, 重要とされる位置についての点の形状と読みが判明した場合でも, 従来の群分類とヲコト点図 (スキーマ) である主要ヲコト点図 26 種の情報を用いて予測されるヲコト点図を表示する, というような処理を行おうとすると問題があると考えられる。例えば, とある資料に基づいて作成された点図において, 四隅の読みが一致し第 2 群点の喜多院点と判断されたが, それ以外の位置の読みについては全く一致せず, システムにおける読みと形状の推薦が正しく機能しないといった事が考えられる。

また 5 章の (2) の, ヲコト点図の編集距離に基づく計算の場合でも 9 割近い点図が築島の群分類と一致する結果となった。一方で, 築島の群分類と異なる結果となった点図のうち, 点の数が

表 5 条件 ii -①で分類できなかった点図

訓点資料名	所蔵	書写年代	加点年代	条件 i -①のとき 類似判定された 26 種ヲコト点	条件 i -① のときの 群分類	条件 i -① のときの 一致数
妙法蓮華經八卷	新薬師寺	奈良時代	平安時代 極初期	古紀伝点	第五群点	1
東大寺諷誦文稿 一卷	佐藤辰次郎 氏所蔵	平安時代 初期	平安時代 初期	妙法院点	第一群点	1
羅摩伽經一卷	聖語蔵	768 年	平安時代 極初期	順暁和尚点	第八群点	1
願經四分律十一 卷	聖語蔵	天平 12 (740) 年	平安時代 極初期	甲点図	第一群点	1
大方廣佛華嚴經 十九卷	聖語蔵	奈良時代	平安時代 極初期	一致なし	一致なし	0

少ないことが影響していると考えられる判定結果となった点図が存在した。例えば、築島の群分類では第 2 群点となっている西大寺所蔵の「金光明最勝王經十卷」では、第 5 群点である遍照寺点との編集距離が 16、第 2 群点である喜多院点との編集距離が 17 となり、僅差で第 5 群点と判定された。「金光明最勝王經十卷」のヲコト点図の星点「・」は四隅にある 4 つの点がハ・テ・ヲ・ニの読みであり、第 2 群点の特徴と一致しており、5 章の(1)では第 2 群点の喜多院点ともっとも近いと判定された。一方で、四隅以外の点では他に近い点図が見つからず、結果として編集コストが低い、比較的星点「・」の少ない遍照寺点と一致する結果となった。このことから、編集距離を用いて従来の群分類と同様の分類を機械的に行う場合には、計算コストについて調整が必要である。一例として、点の記載があり一致しているものに重みを付けるといった方法が考えられる。また一方で、字画の中の点については比較検討の余地があるとも言える。

7. おわりに

本稿では訓点データベースを活用しヲコト点図を機械的に分類する手法について述べた。既存研究で分類されたものを正解とした場合、一致したヲコト点の数に基づく判定方法では、比較の対象とする点の形状を「・」のみとし、ヲコト点がよく付与されるとされる 9 箇所に重みをおいて計算をした場合、159 点図中 153 点図が正解の群に分類できた。また、ヲコト点図を特殊な文字列と捉え、ヲコト点図の編集距離を計算し分類した結果においても、159 点図中 142 点図と 9 割に近い点図を従来と同様の群に分類することができた。また、特殊点と呼ばれる従来の分類では第 1 群～

8 群のいずれにも分類されていなかった点図についても分類を試みた。

実験結果から、点図を作成する元となった訓点資料に関する情報を用いずとも、従来研究における群分類と同様の結果を機械的に得ることが可能であることがわかった。一方で、四隅や字画の中心といった重要とされる位置以外の点を用いた場合、従来研究の群分類とは一致しづらいことがわかった。そのため、四隅や字画の中心の点からヲコト点図（スキーマ）を推定し、それ以外の位置の点について推定を行う手法は有効ではないと考えられる。

今後の課題として、四隅や字画の中心の点以外の位置にある点に関する解析が挙げられる。また、ヲコト点図を特殊な文字列と見立てた編集距離計算の手法についても改善の余地があると考えられる。

参考文献

- [1] 中田祝夫：古点本の国語学的研究（総論編・訳文編），講談社(1954).
- [2] 築島裕：平安時代訓点本論考，汲古書院(1986).
- [3] 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).
- [4] 春日政治：西大寺本金光明最勝王經古点の国語学的研究，斯道文庫(1942).
- [5] 大坪併治：訓点資料の研究，風間書房(1968).
- [6] 築島裕：興福寺本大慈恩寺三蔵法師伝古点の国語学的研究（訳文篇・索引篇・研究篇），東京大学出版会（1965-1967）.
- [7] 太田次男，小林芳規：神田本白氏文集の研究，勉誠社(1982).