

点図情報入力支援ツールによるヲコト点図の電子化

堤 智昭 (東京農工大学 工学府)

田島 孝治 (岐阜工業高等専門学校 電気情報工学科)

高田 智和 (国立国語研究所 理論・構造研究系)

これまで、日本語史の研究資料として訓点資料が利用されてきた。しかし、これらの資料を計算機において統計的に処理できる状態で電子化する手法は十分に確立されていない。本稿では、ヲコト点図を電子化するためのデータ構造を検討する。さらに、ヲコト点図データ構造に従ってヲコト点図の電子化を支援する入力支援ツールを作成し、主要ヲコト点 26 種の電子化を行う。最後に、電子化したヲコト点図データを用い、ヲコト点図間の類似度解析を行い従来のヲコト点の分類と比較する。

Computerization of *Wokototenzu* for Digital Analysis with an Input Support Tool

Tomoaki TSUTSUMI (Tokyo University of Agriculture and Technology)

Koji TAJIMA (National Institute of Technology, Gifu College)

Tomokazu TAKADA (National Institute for Japanese Language and Linguistics)

Kunten materials are used as a research material of Japanese etymology. However, the method to computerize in the state with which *Kunten* materials can be treated statistically in a computer is not established sufficiently. Therefore in this paper, we discuss the method to computerize the *Wokototenzu*. *Wokototen* is one of *Kunten*. And *Wokototenzu* is illustration of *Wokototen*. We make the tool which computerizes the *Wokototenzu*. Finally we do a similarity analysis during the *Wokototenzu*.

1.はじめに

日本語史の研究では、訓点資料を利用した言語研究が行われている。訓点資料とは、漢籍、仏典、国書などの漢文を日本語として理解し読むために、当時の読者により訓点に加えられた資料である。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。平安・鎌倉時代の訓点資料には、ヲコト点と呼ばれる特徴的な訓点が付与されている。ヲコト点は、星点、線点、鉤点といった多様な記号を用い、助詞や活用語尾などを表す訓点である。

また、一つの漢文に対して墨点、白点、朱点、角点など複数の体裁(色)で点が付与されている訓点資料も確認されている。これは、この漢文資料を複数の読者が読み、それぞれが解釈して訓点を付与したためと考えられている。このことから、異なる体裁の訓点を比較して時代の変化を読み取る資料として、活用されている[1]。

これまでに筆者らは、ヲコト点を含む漢文訓点資料を電子化するための構造化記述手法を検討してきた [2][3]。ヲコト点を付与した電子データの作成を試みるため、中心を(0, 0)とする7×7マスのグリッド座標を定義し、電子化作業を支援するツールを試作した。

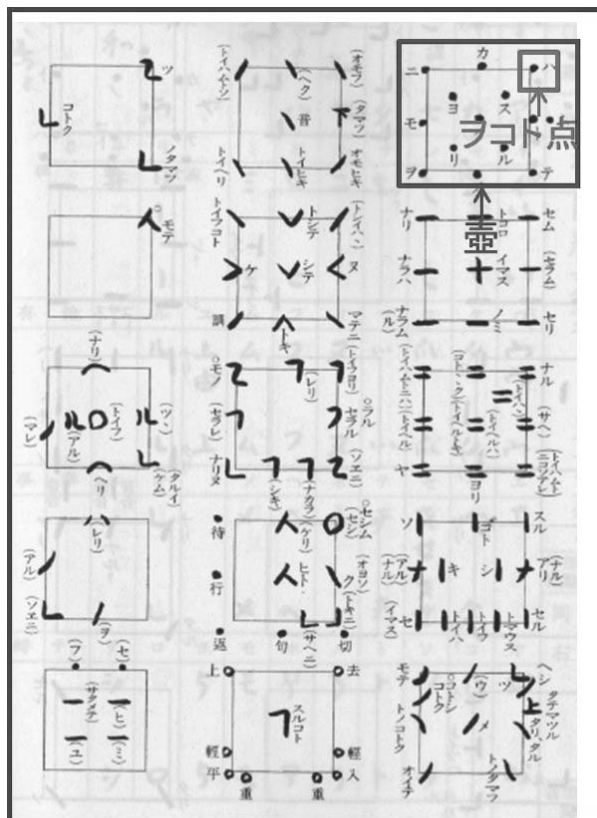
本稿では、ヲコト点図を電子化するためのデータ構造を検討する。また、検討した構造化記述方

式に従ってヲコト点図データを作成する支援ツールを試作し、それを用いてデータ入力を行う。入力するヲコト点図には、文献[4]に記載された主要ヲコト点 26 種を用いる。最後に、電子化した主要ヲコト点 26 種の点図データからヲコト点図間の類似度解析を行い従来のヲコト点の分類と比較する。

2.ヲコト点図の特徴

ヲコト点は、漢字の字画の四隅や内部、周辺に付与する記号であり、原文のどの位置にどの形状の点が増えられるかによって意味が異なる。増えられる記号は「・」「ー」「。」等多岐にわたり、一般的な計算機が表現できる字系には存在しない記号も使われている。その配置の形式は、使用者の学派や時代によって異なり、東大寺点や喜多院点など名前がつけられている。現在までに確認されているヲコト点の種類は 100 種を超えている。中田祝夫[5]は、代表的なヲコト点 26 種類をヲコト点の歴史の変遷・発達を考慮し 8 つの群(第 1 群、第 2 群点、第 3 群点・・・第 8 群点)に分類している。

同じ形状でも、増えられる点によって意味が異なる一例として、石山寺・天理図書館所蔵の金剛般若経集験記巻第二における星点(●)を例にとり説明する。金剛般若経集験記巻第二では、星点(●)が漢字の中央上に加点されていたら「ヲ」、



↑
点図

図 1 ヲコト点図例 (喜多院点) [4]より引用

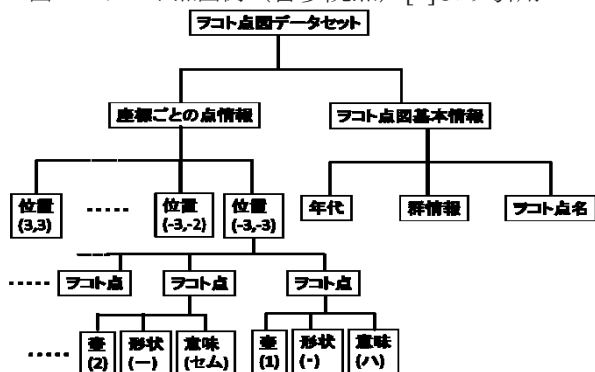


図 2 ヲコト点図データ構造概要

漢字の右上ならば「ト」といった意味になる。また、同じ形状、同じ位置にある点でもヲコト点の種類が異なれば対応する意味も異なる例としては、先ほどと同じ形状の星点(★)が漢字の右上に加点されていても「ハ」の意味になる、金剛般若経集験記巻第二と国立国会図書館所蔵の大毘盧遮那成仏神変加持経巻第六との違いが挙げられる。

ヲコト点図とは、ヲコト点の形状と意味、付与された位置の組み合わせから、そのヲコト点がどのような意味を持つかを図示したものである。ヲ

表 1 XML タグの定義

XML 要素名	内容
群番号	ヲコト点図の群番号を表す
ヲコト点名	ヲコト点の名前を表す
年代	ヲコト点が使われたとされる年代を表す
data	ヲコト点図データの親要素
grid	1 マスのグリッドを表す, data タグの子要素
位置	グリッドの位置情報を表す, grid タグの子要素
X	X 軸方向の位置情報を表す, 位置タグの子要素
Y	Y 軸方向の位置情報を表す, 位置タグの子要素
ヲコト点	1 つのヲコト点を表す, grid タグの子要素
形状	ヲコト点の形状を表す, ヲコト点タグの子要素
意味	ヲコト点の意味を表す, ヲコト点タグの子要素
壺	ヲコト点の壺を表す, ヲコト点タグの子要素

コト点図では、ヲコト点を星点、線点、鉤点など形状によって分類し、同じ形状を持つヲコト点を図 1 のように一つにまとめて位置と読みと共に図示する。このヲコト点の形状による集合を壺と称する。ヲコト点図は、壺の集合として定義される。

3. ヲコト点図電子データの構造

ヲコト点図の電子化を行うにあたり、出力する電子データは図 2 に示す木構造のデータ構造とした。主要ヲコト点の点図を一つの単位としてデータセットを作る。一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 マスのグリッド座標の位置情報を親とし、点の形状と意味、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。

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 1 の通りとした。

4. ヲコト点図入力支援ツールの設計と実装

本ツールは、ヲコト点図電子データのデータ構造にしたがって、ヲコト点図の電子化作業を支援

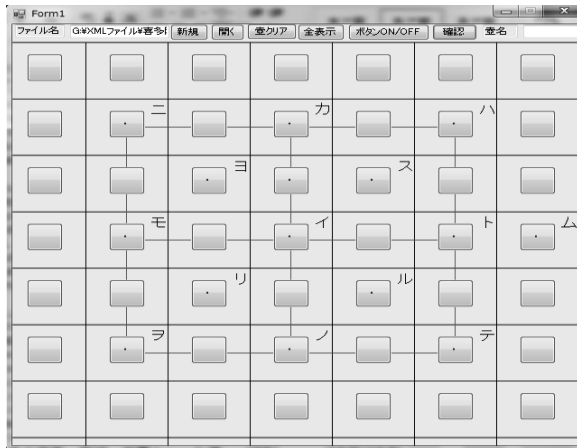


図 3 支援ツールメイン画面

するツールである。図 3 に示すようなヲコト点図の資料から、XML 形式のテキストデータを作成する。本ツールを用いることで、データ入力者は、直感的な操作でヲコト点図を電子化可能である。実装には言語 C# を用いており、.NetFramework4.0 が導入された計算機であれば OS を問わず動作する。

4.1 GUI 設計

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. XML 確認画面
3. 点情報確認画面
4. 点図データ入力画面

1~3 の画面は、プログラム起動時に表示される。4 の画面は、メイン画面から後述する点図入力処理を行う時に表示される。以下各画面の設計について論ずる。

4.1.1 メイン画面

本ツールのメイン画面の例を図 4 に示す。ヲコト点図集の資料では、図 3 に示すように壺ごとに情報がまとめられている。これを電子化するため入力する場合、資料と同じような形で入力位置や入力結果を表現できると効率が良いと考える。そこでメイン画面では、7×7 マスのグリッドを描画する。さらに点図集の四角形を表現するため中心の 5×5 マスに赤い四角形を描画することで、資料の壺と同じ形を表現した。各グリッドにはクリックボタンを配置し、これをクリックすると 4 の点図データ入力画面を表示する。入力したヲコト点の形状はクリックボタン内に、意味はグリッドの右上にそれぞれ表示される。これにより、入力中の点図データを壺ごとに視覚的に確認しやすい形で表示することができる。また、画面の上部にツールバーを実装し、ファイルの新規作成や既存のファイルを開く等の動作を割り当てた。



図 4 XML 確認画面

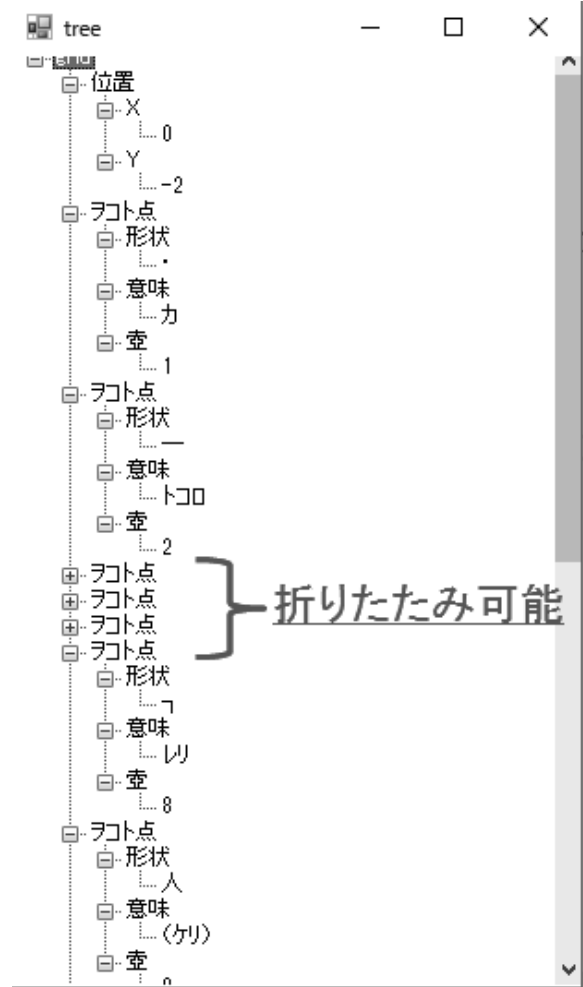


図 5 点図情報確認画面

4.1.2 XML 確認画面

XML 確認画面の例を図 5 に示す。この画面は、出力結果の XML テキストを表示する画面である。メイン画面では視覚的な表現を優先したため、入

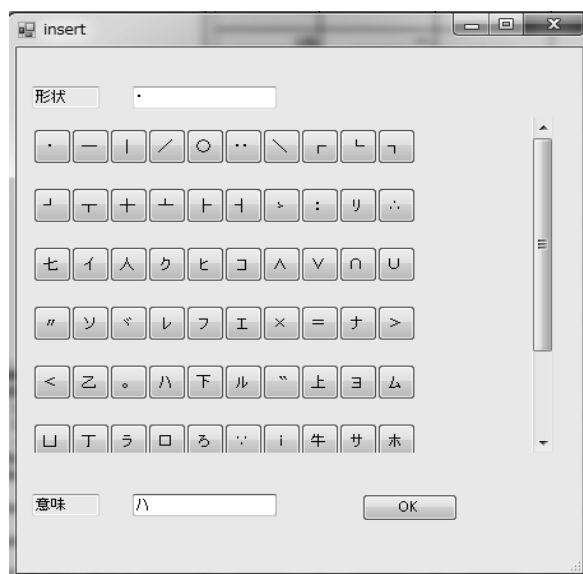


図 6 ヲコト点データ入力画面

力作業中の XML テキストがどのように出力されるか確認できない。そこで XML 確認画面において、ツールが出力する XML テキストを確認可能とする設計とした。他の画面において点図データが変更されると、すぐに XML 確認画面のテキストにも反映される。

4.1.3 点図情報確認画面

点図情報確認画面の例を図 6 に示す。この画面は、グリッドごとに入力されている点図情報を表示する画面である。メイン画面において各グリッドのボタン以外のフィールドをクリックすると、点図情報確認画面にクリックしたグリッドのヲコト点一覧が表示される。また、メイン画面上部の「全表示」ボタンをクリックすると、全てのグリッドのヲコト点一覧が表示される。

表示方法は、グリッド 1 つを親要素とし、その位置情報(X,Y)と入力されているヲコト点情報(形状・意味・壺)を子要素とする木構造とした。各要素は、要素名の左に表示されている□ボタンをクリックすると折りたたむことができる。

4.1.4 ヲコト点データ入力画面

ヲコト点データ入力画面の例を図 7 に示す。ヲコト点データ入力画面では、ヲコト点の形状・意味を入力する。ヲコト点の形状はテキストボックスに直接入力することもできるが、ウィンドウの真ん中に表示されているボタンの中から選ぶこともできる。ヲコト点の形状は既存の Unicode の文字集合にある文字を利用する。例えば「・」は「・」、「一」は「一」などとし、あらかじめ外部のテキストファイルに定義する設計とした。

4.2 データ入力手順

本ツールを用いてヲコト点図データを入力する手順は以下の通りとした。

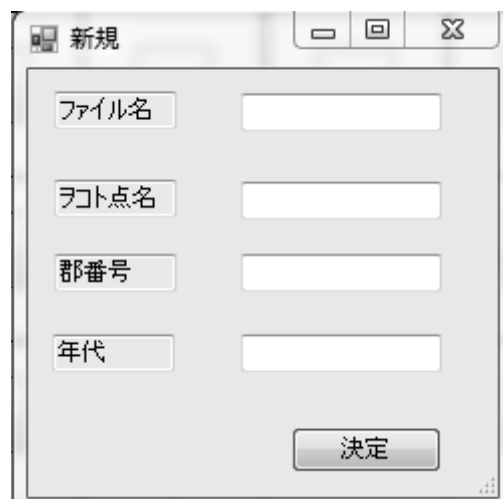


図 7 新規作成画面

(1) ファイルを開く

新規に XML テキストを作成する場合は、メイン画面上部の「新規」ボタンをクリックする。その場合、図 8 に示す新規作成画面が表示される。この画面では、ファイルの名前と、ヲコト点図の基本情報(ヲコト点名、群番号、年代)を入力する。既存のファイルを編集する場合は、メイン画面上部の「開く」ボタンをクリックする。

(2) ヲコト点データ入力

ヲコト点図の入力は、図 1 に示した点図集を見ながら行うことを想定している。そのためデータ入力は 1 壺単位で行うこととした。現在入力している壺の名前は、図 3 のメイン画面右上の壺名テキストボックスに表示される。データ入力するグリッドのボタンをクリックし、ヲコト点データ入力画面を表示し、形状と意味を入力する。例えば、図 1 の右上の点を 1 壺目の形状「・」意味「ハ」として入力する場合、メイン画面壺名に「1」と入力し、赤線の四角の右上 (2,-2) のグリッドのボタンをクリックして形状・意味を入力する。

1 壺分のデータ入力が終わり次の壺を入力する場合、メイン画面上部の「壺クリア」ボタンをクリックする。そうすると、XML テキストを保存したまま、メイン画面の壺名とヲコト点の形状・意味の表示が消去され、新しい壺を入力可能となる。

(3) 入力データの削除

ヲコト点情報を削除する場合、意味のみ、または形状のみ削除するといった場合は考えにくい。そこで、点図情報確認画面を用いて、1 ヲコト点単位で削除を行う設計とした。まず、訂正を行いたいデータを入力したグリッドのフィールドをクリックし、点図情報確認画面に入力済みのヲコト点一覧を表示する。その中から訂正対象のヲコト点要素を選択し、キーボードの Delete キーを押し削除する。

5. ヲコト点図入力

本ツールを用いて、実際に文献[4]に記載された主要ヲコト点 26 種の電子化を行った。図 1 の赤枠で囲った星点の壺を、本ツールを用いて入力すると、図 3 のように表示される。図 1 と図 3 を比較すると、ほぼ同様の配置で点の形状と意味が表示されており、視覚的な入力補助が可能であると言える。

点図データの形状を XML テキストに出力する


にあたって、 のように Unicode の文字集合内の文字では代用できない形状の点が存在した。そこで今回これらの文字は「二母(縦)」のように、二文字以上の組み合わせと、組み合わせる方向(縦、横)をカッコ書きで表す方式で代用した。主要ヲコト点 26 種を全て電子化した結果、表 2 に示す 6 つの記号の代用が必要であった。

表 2 代用した記号一覧 (形状は[4]より引用)

ヲコト点図中の形状	代用した組み合わせ
	一 一 (よこ)
	ケ人 (よこ)
	二七 (たて)
	二母 (たて)
	一 一 (たて)
	一 一 (たて)

6. 点図間の類似度解析

6.1 ヲコト点図の類似度計算方法

本ツールを用いて電子化したデータの応用例として、電子化した主要ヲコト点 26 種の点図データからヲコト点図間の類似度解析を行った。ヲコト点図データは、複数の壺の集合であり、壺は複数のヲコト点から構成されている。そこで 2 つの点図間の類似度を計算するには、2 点間に存在するヲコト点が、それぞれどの程度似ているかを計算し、その合計を求めれば良いと考えた。

今回対象とした主要ヲコト点 26 種の点図データでは、それぞれの点図ごとに記載されているヲコト点の数が異なる。そこで、異なる長さの文字列に対して文字列の距離を計算する時に用いられる、レーベンシュタイン距離計算のように、一方のヲコト点をもう一方のヲコト点に変形するのに必要な、点の「移動」、「挿入」、「削除」といった操作の最小回数を基準として類似度を点数化した。また、ヲコト点図データセットでは、7×7 マスのグリッド座標と点情報が記述されている。一般的に、座標軸上に展開されたデータは、2 つの要素同士の距離を測ることでその要素間の類

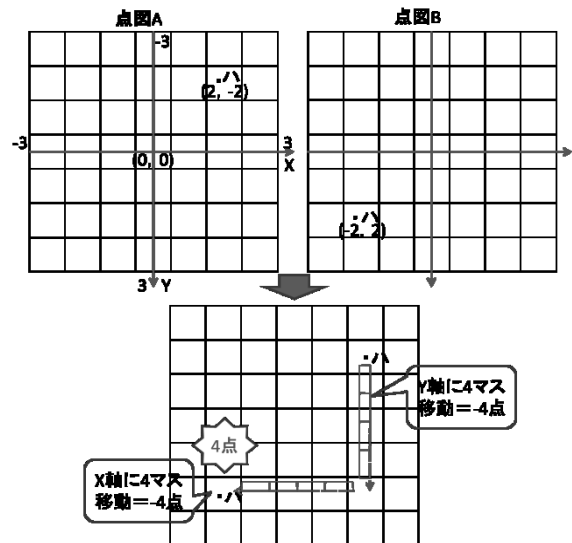


図 8 類似度の計算方法例

似度を求めることができ、距離が近いほど類似度が高いと考えられる。そこで、点の「移動」操作に重みをつけ、2 点間の距離が近いほど類似度が高くなる計算方式を用いた。また、ヲコト点の位置は整数のグリッド座標を用いて表現されているため、2 点間の距離計算には X 軸、Y 軸に分解して簡単に計算可能なマンハッタン距離を用いた。

具体的な類似度の計算方法は以下のとおりとした。

1. 点の形状と意味が同一のものが存在すれば 12 ポイント加点する ((-3,-3)から(3,3)まで移動すると 0 点となる配点)
2. 点の位置が X 軸 Y 軸共に一致すれば減点は無しとする
3. 点の位置が X, Y 軸各方向に 1 マスずれるたびに -1 ポイントとする
4. 点の形状と意味が同一のものが存在しなければ加点しない (挿入、削除が必要な場合は加点無し)

例えば、図 8 のように点図 A の星点の壺(2,-2)の位置に形状「・」意味「ハ」の点があり、点図 B の星点の壺(-2, 2)の位置にも同一の形状意味の点が合った場合、それぞれの点の位置は Y 軸に 4 マス、X 軸に 4 マスずれているため、 $12 - 4 - 4 = 4$ ポイントとなる。

6.2 類似度計算結果

6.1 の計算方法に従って、主要ヲコト点 26 種の星点 (・) の点図データについて類似度の計算を行った。計算対象を星点のみとしたのは、星点は 26 種全てに存在し、これまでのヲコト点の研究においても星点を基準に分類されてきたため、従来のヲコト点分類論と比較することができる

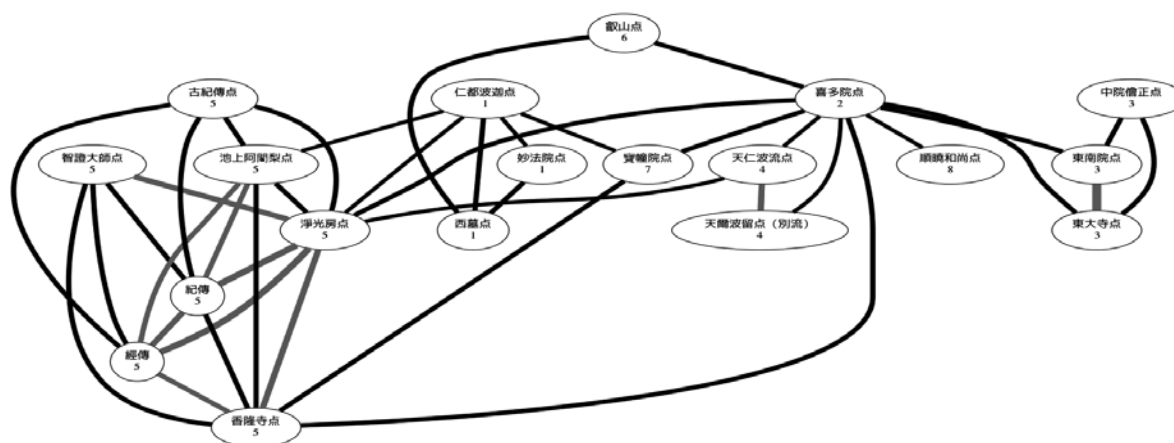


図 9 点図間の関係性

表 3 類似度計算結果 (上位 10 通り)

点図の組み合わせ	群番号	類似度
東南院点-東大寺点	3 群 - 3 群	156
浄光房点-経傳	5 群 - 5 群	128
浄光房点-紀傳	5 群 - 5 群	128
浄光房点-香隆寺点	5 群 - 5 群	120
天仁波流点-天爾波留点 (別流)	4 群 - 4 群	119
紀傳-経傳	5 群 - 5 群	118
智證大師点-浄光房点	5 群 - 5 群	112
地上阿闍梨点-経傳	5 群 - 5 群	108
経傳-香隆寺点	5 群 - 5 群	108
地上阿闍梨点-紀傳	5 群 - 5 群	107

と考えたためである[4]。点図同士の組み合わせは ${}_{26}C_2$ (325) 通りである。表 3 に類似度の高かった点図の組み合わせ上位 10 通りを示す。また、図 9 に類似度の計算結果を用いて点図間の類似度が高かった組み合わせ 40 通りを線で結び、その関係性を図示する。点図名の下に数字は、その点図が所属する群番号を示す。2 点図間を結ぶ線は、太いほど類似度が高いことを示す。また類似度の上位 10 点図間は赤線で結んだ。描画には Graphviz を用いた。

結果から、第 5 群点を中心として、同一の群に所属するヲコト点同士の類似度が高いことがわかる。また第 1 群点の派生と考えられている第 5 群点と第 1 群点間の類似度が高いことがわかる。一方、従来は第 5 群点と結びつきが強いと考えられていた第 6 群点の叡山点は、第 1 群点の西墓点や、第 2 群点の喜多院点との類似度が高いという結果が得られた。これは、従来の分類では主に星点の 4 隅の違いを基準に分類していたのに対し、本方式では星点の全ての点を平等に比較したためではないかと考えられる。

7.まとめ

今回、計算機による訓点情報の比較を目的とし、

ヲコト点図を電子化するためのデータ構造の検討、ヲコト点図電子化支援ツールの試作、主要ヲコト点 26 種の電子化を行った。点図情報のデータ構造は、7×7 マスのグリッド座標の位置情報を親、点の形状、意味、壺を子とする木構造とし、電子ファイルは XML 形式で記述した。また、電子化した点図データを用いて、ヲコト点図間の類似度の計算を行った。

今後の課題として、星点以外の壺を用いた比較、類似度に応じたヲコト点のクラスタリングと従来の 8 群分類との比較等、電子化データを用いたさらなる解析、ヲコト点専門家も交えた類似度の計算手法の検討が挙げられる。

謝辞

本研究は JSPS 科研費 15H06833 の助成を受けたものです。

本研究は国立国語研究所共同研究プロジェクト「文字環境のモデル化と社会言語科学への応用」による研究成果の一部です。

参考文献

- [1] 小助川 貞次：訓点資料の展開史における有鄰館蔵『春秋経伝集解卷第二』の位置(<特集>資料研究の現在), 日本語の研究 Vol.4, No.1, pp.15-30(2008)
- [2] 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 巻 7 号, pp.211-216(2012).
- [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 (June 2013) .
- [4] 築島裕：訓点語彙集成<第 1 巻>, ヲコト点概要, 汲古書院(2007).
- [5] 中田祝夫：古点本の国語学的研究, 講談社 (1954) .