# 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

田島　孝治* 堤　智昭★ 高田 智和†
*岐阜工業高等専門学校　電気情報工学科 ★東京農工大学　電子情報工学専攻
†国立国語研究所 理論・構造研究系

本稿は，漢文訓読に使われる訓点であるヲコト点を電子的に扱うためのデータ構造を提案する．提案方式の特徴は，漢文中の一文字に対し，文字の中心を原点とする７×７の座標系を作り，付与されたヲコト点の座標を定め，文字を親，ヲコト点を子とする木構造のデータ形式で保持することである．今回はさらに，定義したデータ構造を用いて，ヲコト点が付与された資料を電子化し，表示，訓読するための入力支援システムを試作したので，その機能についても述べる．また，システムの実装の結果明らかになった問題点についても議論する．

# A Data Stricture and Input Support Tools for the Diacritical Mark(*wokototen*) in the Glossed Materials

TAJIMA Koji† TSUTSUMI Tomoaki★ TAKADA Tomokazu‡
†Gifu National College of Technology
★Tokyo University of Agriculture and Technology
‡National Institute for Japanese Language and Linguistics

We propose a data structure and input support tools for the documents marked up with the diacritical marks(*wokototen*). We discuss the problem about the diacritical mark processing on the computer. We define the integer coordinate system with its origin based at the center of the character to structurize the diacritical marks. And we make two tools for support to input the documents marked up with the diacritical marks. One tool can make a XML text used the defined integer system, and the other tool can replace a XML text to a transcription in Japanese.

## １．はじめに

漢文訓読とは，漢文（主として中国語文）で書かれた原文に句読点や返点，訓などを付与し，読者の言語として解読できるようにすることである．漢文をそのまま理解し，読むことができない読者でも，訓読法の規則を覚えれば，中国大陸から届く文献を読むことができるため，日本では奈良時代以降広く利用されてきた．訓読符号も歴史的変遷を遂げているが，平安・鎌倉時代を中心に広く使われていたものに，ヲコト点がある．

ヲコト点は，原文の漢字に重ねて点を打ち，その形状と位置によって助詞や語形の一部に相当する音節を表す訓読符号である．形状・位置と読みの対応は，時代や使用者の学派によって異なっており，100 種類以上のヲコト点が発見されている[1].

ヲコト点は，墨筆による墨点の記述だけではなく，白点，朱点，角点などの体裁がある．また，一つの漢文に対して，複数の体裁でヲコト点が付与されている場合もある．これは，同じ写本や刊本を異なる読者が読んだ場合などに現れることが多い．内容に対する解釈の違いから訓点が追加，削除されたり，時代の変化によって，対応が異なる訓点が施されたりすると，複数の体裁が表れる．

現在，漢文加点資料の電子化においては一般的にヲコト点を記録するのではなく，原文のみ，または，ヲコト点を解釈した書下し文での記録が多く行われている．この方法は，原文を読みやすく記録できる反面，解読において他の解釈を行うことが難しい．また，漢文訓読に関する専門的な知識がなければデータ化することができないという問題もある．さらに，ヲコト点の情報は残っていないため，ヲコト点自体の分析行う資料として利用できない問題もある．

## ２．目的

本稿では，これらの問題を解決するためにヲコト点を含む漢文加点資料を電子化するための構造化記述手法を検討する．また，ヲコト点情報の付与と書下し文の生成を行うツール試作したので，これを報告する．

ヲコト点を含む資料の電子化の流れを図１に示す．初めに資料からテキストデータを作成する．これはプレーンテキストであり，純粋に漢文のみを記録する．次に，この漢文テキストに対し，ヲコト点情報を追加し，構造化したテキストデータを作成する．一次資料における文字

単位で，文字に付与されていたヲコト点の体裁や位置を記録する．電子データ形式の体裁を一意にするために，入力作業は専用のツールを製作して行う．

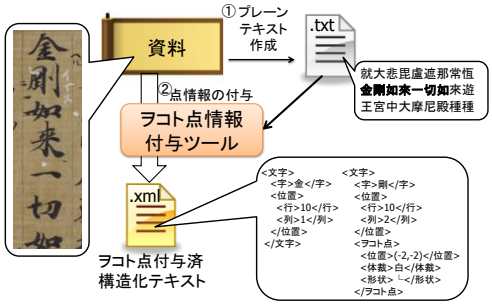



**図１　ヲコト点情報を付与した構造化データの作成**

ヲコト点情報を付与した構造化テキストは，専用のソフトを使うことで書下し文に変換できるようにする．変換時には，あらかじめ作成しておいた，ヲコト点と音節の対応データ（以後，点図データと呼ぶ）を用いる．これにより，点図データを複数用意することで，訓読結果を任意の点図に切り替えることを可能にする．また，書下し文を作るだけではなく，Web ブラウザ等でヲコト点を目に見える形で再現できるようにすることも検討している．

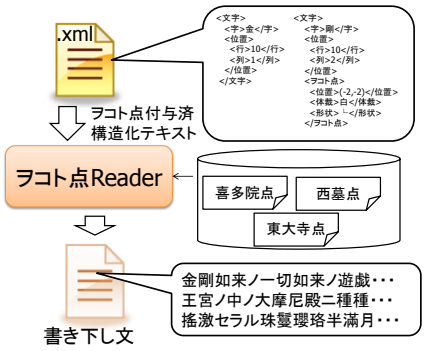



**図２　ヲコト点付与済みテキストの活用方法**

## ３．ヲコト点の構造化記述に関する議論

ヲコト点の情報を電子的なデータファイルとして記録するには，点の体裁と，形状，位置をどのように記述するかが課題となる．

点の体裁は，墨，朱，白，角のようにテキスト形式で記述可能である．文献によっては朱や墨が２色以上使われているものもある．しかし，同じ色で３種類以上の記述が行われることは少ないと判断し，「朱１」，「朱２」のように記述することにした．

次に点の位置と形状に関して検討する．図３のように，ヲコト点は点の形状（この場合は星点）と位置により，対応する音節が決まっている．しかし，対応する音節は，位置と形状だけで決まるわけではない．例えば文字の左上の星点が，喜多院点では「ハ」，西墓点では「ニ」の意味となる．このように，点の意味は，点の位置だけでなく，ヲコト点の系列によって異なり，位置だけ記録しても，ヲコト点の系列が分からなければ，読むことはできない．また，単に「右上」と表記しただけでは，文字から離れているのか，重なる位置なのかが分からない．

これらの理由から，書下し文を作るためだけであれば，位置，形状，ヲコト点の系列から意味を解釈して，ヲコト点の情報は「ハ」とだけ記録するほうが都合が良い．一方で，ヲコト点の時代的な変化の分析や，どの系列か分からないヲコト点の分析には，点の位置を絶対的な座標として記録する必要がある．

これまでに，ヲコト点と同様に点や記号を使った漢文用の訓点である「点吐口訣」を表すための，左上を原点とする５×５の座標系が提案されている[2]．2000 年に韓国で角筆資料が発見されて以降，韓国角筆の記述と解読，分析のために，韓国角筆資料の研究者の間で広く利用されている座標系である．

この座標系で日本のヲコト点を表そうとした場合，小数を使わなければ表せない点図がある．小数を用いると，一次資料中の訓点の位置が少しずれていた場合の記述方法が一定でなくなる可能性がある．すると，音節（読み）と点の位置とを対応させる際に，関係が１：１でなくなるという問題が発生する．また，ヲコト点は時代の変化により，その位置が回転することが分かっており，回転処理を行いやすい座標系が適していると考えられる．

そこで，ヲコト点を表すのに適した新たな座標系を考える．座標系の設計においては，(1) ヲコト点の重要な点は文字の四隅と文字の中心である．(2) 四隅に打った点が表す音節は，時代の変化とともに，回転してきた経緯がある．という２点に注目した．

今回設計したヲコト点の位置を表すための，座標系を図４に示す．座標系は，資料中の絶対的な位置を表す座標ではなく，一つの文字に対して決定される相対座標系とする．さらに，点図集[1]に示された文字の内側，縁，外側のヲコト点を全て整数で表すために，７×７マスのグリッドを作る．また，中心点の重みが強く，回転処理を容易とするため，原点を中心とした．

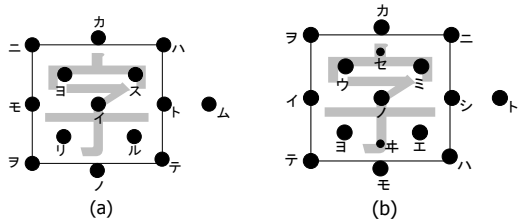



**図３　喜多院点(a)と西墓点(b)の点図**

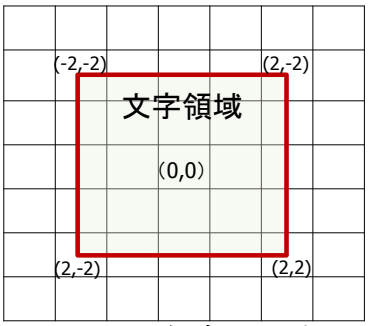



図４　ヲコト点データの座標系

## ４．入力支援システムの設計と試作

### 4.1　入出力データの仕様

本システムは，「ヲコト点情報付与ツール」と「ヲコト点 Reader」という二つのツールで構成される．「ヲコト点情報付与ツール」に入力するデータは，手動で入力したプレーンテキストである．また，出力するデータはヲコト点情報を付与した構造化テキストである．一方，「ヲコト点 Reader」は，「ヲコト点情報付与ツール」で生成した，ヲコト点情報を付与した構造化テキストと，点図データを読み込み，書下し文のプレーンテキストを出力する．次にこれらのデータファイルの仕様を述べる．

(1) 漢文のプレーンテキスト

手作業で作成するプレーンテキストは，基本的に漢文のみを記録するデータである．このデータの作成は資料（原本・複製・影印など）を見ながら行うことになる．漢字のみを入力するだけでも問題ないが，ヲコト点を付与する作業においては，対象の文字が，資料中の物理的にどこにあったかを表す行（以後，物理行と呼ぶ）を記録しておいたほうが望ましい．さらに，資料を作る際に墨筆や朱筆で補入や見せ消ちが行われている場合もあり，これらが別の観点でこの資料を分析する際に役に立つ可能性は高い．また，プレーンテキストには入力できない文字（外字）が現れる場合もある．これらは資料を見ながらでなければ入力することができず，必要になったときにもう一度資料を見直して入力しなおすことは効率的ではない．そこで，表１のような簡易的なタグをプレーンテキスト中に埋め込むことにする．

(2) ヲコト点情報付与済み構造化テキスト

ヲコト点情報を付与したテキストデータの表現には，XML 形式を採用する．ヲコト点は一つの文字に対して，複数の点が従属する関係を持つため，木構造での表現が適している．XML 形式は汎用性の高いテキスト形式であり，木構造を表現できるため，今回はこれを採用した．

表１　プレーンテキスト用の簡易タグ

| 区分 | 種別 | フォーマット |
|---|---|---|
| 行 | 物理行 | 行番号:本文 |
| 符号 | 割行 | 〈本文〉 |
| | 割行内改行 | / |
| | 墨筆補入 | +b{本文} |
| | 白筆補入 | +w{本文} |
| | 朱筆補入 | +r{本文} |
| | 墨筆見消 | $b{本文} |
| | 後補部分 | #a{本文} |
| 外字 | 諸橋番号 | =[Mxxxxx] |
| | UCS | =[U+zzzz] |

XML の構造を図５に示す．本文を文字の集合とし，各文字に対し，(A)字，(B)資料上における行，列，(C)ヲコト点を付与する．ヲコト点は一文字に対し複数付与されていても構わない．本データの構造は，「文」→「文字」→「ヲコト点」という階層構造であるが，文字間をつなぐ「合符」のような訓点も存在するため，複数の「文字」要素を子に持つことが可能な要素を追加する予定である．ヲコト点は，３節で述べたとおり，体裁，形状，位置を要素として持つことにする．

XML 中でタグを表す文字列に関しては，日本語，英語等いくつかの表記が考えられるが，データ構造に影響を与えるものではないため，試作段階においてはタグ名を日本語で記述することにした．

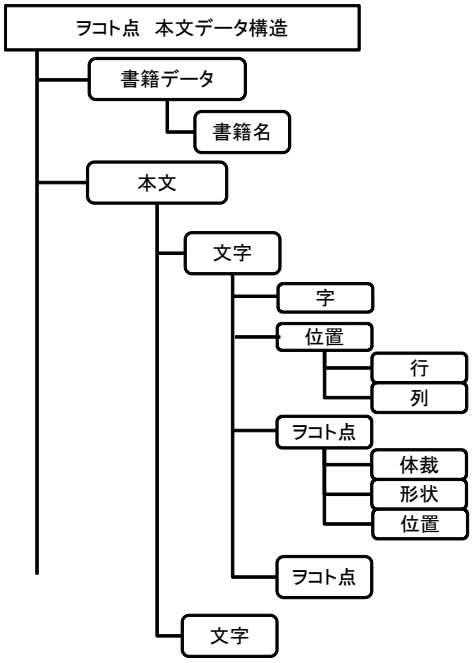



図5　ヲコト点情報付与済みテキストの構造

(3) 点図データ

書下し文作成のために利用するヲコト点図の点の位置と対応する音節（読み）をまとめたデータである．電子ファイルは XML 形式とし，点図集[1]に記載された情報を表現できるデータ構造を検討した結果，図 6 に示す構造とした．何を中心にするかにより構造が異なるが，変換処理における点から音節の検索を行う際に，検索キーを数字にできるほうが効率的と考え，整数座標である位置を親とした．この結果，同じ位置座標を持つ点を Grid としてまとめた．また，同じ記号を持つヲコト点を「壺」の要素でグループ化できるようにした．

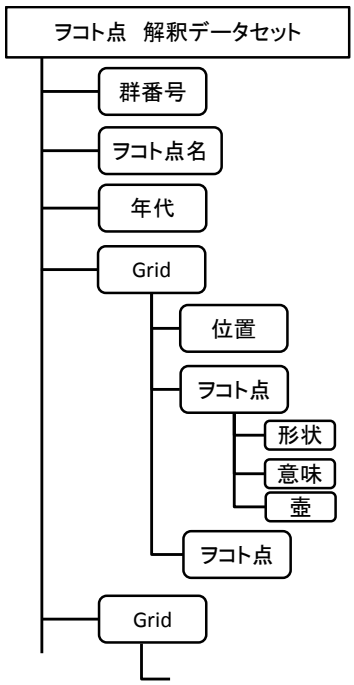



図6　点図情報保存用 XML の構造

4.2　ヲコト点情報付与ツール

本ツールは資料からヲコト点を読み取る作業（移点）を行う際に，プレーンテキストに対しヲコト点情報を加え構造化テキストを作成する処理を行う支援ツールである．システムの GUI と使い方を図 7 に示す．本ツールは，簡易タグを解釈しながらプレーンテキストの漢文データを読み込み，ヲコト点情報を付与した後は XML ファイルとして出力する．

プレーンテキストを開くと，テキスト中の文字が縦書きで表示される．ここで，ヲコト点を付与したい文字をクリックすると，点情報を入力するためのダイアログが開く．利用者は，資料を見ながら，点の体裁と形状を選択し，位置を示すグリッドをクリックすると，文字に点情報を追加できる．

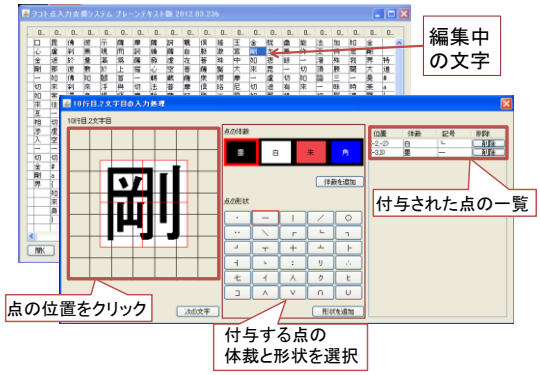



図７　ヲコト点入力支援ツール

作業は 1 文字単位で行い，途中結果も保存できるようにした．途中結果も含め，保存形式はすべて XML ファイルである．途中結果だけでなく，すべて入力が終わったデータも読み込むことができるので，簡易的なヲコト点ビュアとしても利用することができる．

4.3　ヲコト点 Reader

このツールは，ヲコト点情報付与ツールを利用して作った XML データから，書下し文を機械的に生成する機能を持つ．ツールにヲコト点情報付与済み構造化テキストと点図データを読み込ませると，ヲコト点が持つ，点の形状，位置，音節（読み）の関係を対応させて，書下し文が作られ，漢文を訓読できる．

ツールの GUI を図 8 に示す．右上のドロップダウンメニューから任意の点図データを選択すると，書下し文が機械的に生成される．点の解釈が正しく，適切な点図を選択すれば，加点当時の解釈に基づいた書下し文が生成されるはずである．しかし，実際には，利用者による点の位置の記録ミスや，文献によるヲコト点図の差異によって，部分的に適切でない書下し文が出力されると考えられる．

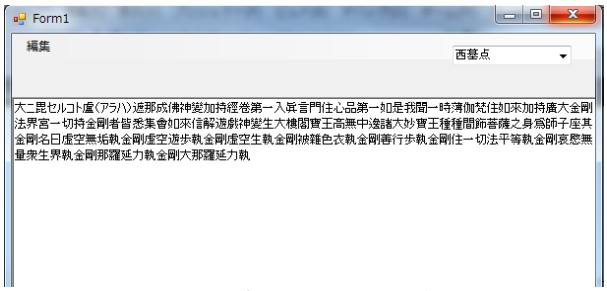



図８ ヲコト点 Reader の動作画面

本ツールは，適切でない書下し文が出力されるヲコト点を持った資料を，様々な点図データを当てはめながら解析していく際に利用することを想定している．このため，ツールが利用可能なヲコト点図は，文献[1]に記載された，喜多

院点をはじめとする 26 種類だけでなく，点図を原点中心に 90 度単位で回転させたものも扱える．これにより，利用者はのべ 104 種類の点図を当てはめて，漢文に付与されたヲコト点を分析することができる．

さらに，文献によっては，墨筆は喜多院点，角筆は西墓点ということもあると考えられるので，点の体裁ごとに書下し文を作れるようにしている．もちろん，墨筆の訓点に朱筆を書き加えたというケースを考慮して，二つの体裁の訓点の合成結果を組み合わせることもできる．

なお，本ツールで利用する点図データの種類は多く，XML を手作業で入力すると，入力ミスが発生する可能性が高い．そこで，点図データを作成する専用ツールも開発した．このツールを使えば，新たなヲコト点図が必要になっても，簡単に追加することができる．

## 5．検討課題

本ツールの動作検証として，国立国語研究所が試験公開している『金剛頂一切如來眞實攝大乘現證大教王經』を資料として入力作業を行った．入力したデータは一部分だけであるが，作業の結果，いくつかの問題と検討すべき点が明らかとなった．なお，この写本は喜多院点（図 3 (a)）のヲコト点が主に白点で付与されている．

### 5.1 文字の形と点の位置の関係

ヲコト点情報付与ツールにおける，点情報の入力作業は，まず，利用者が文字を選択してグリッドを表示させ，さらにグリッド内で点の位置をクリックする方式である．ここで問題となるのが，文字の形によって，点の位置が変化することである．図 7 の「剛」のような，輪郭が四角形に近い文字であれば，この問題は起こらない．問題が起こるのは「一」，「入」，「上」，「中」のような，形状が極端に横長の文字または，三角形の文字である．

図 9 に今回の作業の中で出現した，四角形でない文字に付与されたヲコト点の例を示す．(a)は，点がどちらに付与されているか分かりにくいと思われる例である．一見すると，「界」の左下の星点か，「令」の左上の星点かの区別がつきにくい．一方で，(b)の点は「无（無）」中央下と見える可能性は少ないが，「上」の中央に星点が付与されているように見える．最後の(c)の点は「<」と「下」の点が付与されているが，この二つはどちらも文字の右側に記述するヲコト点である．

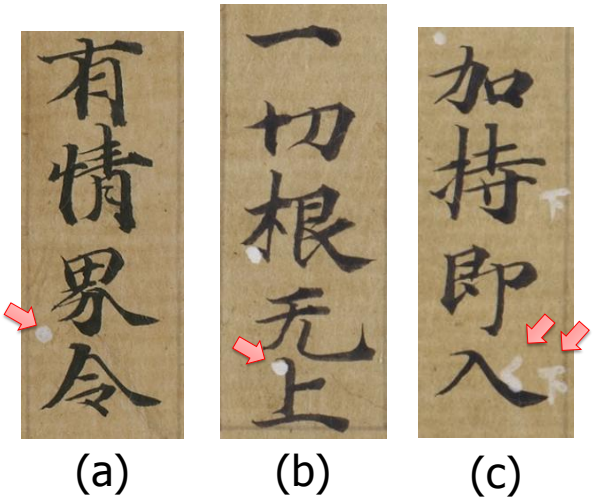


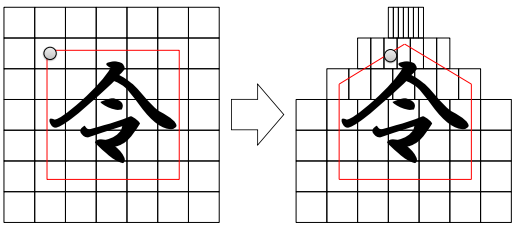

**図 9　文字の形によるヲコト点の位置のずれ（金剛頂一切如來眞實攝大乘現證大教王經より）**

これらの解釈は，ヲコト点が加点された訓点資料を読んだ経験の有無によって異なってくると思われる．(a)の場合，星点は「界」についていると考えるのが普通である．また，(b)の場合の星点は「上」の左上の点である．

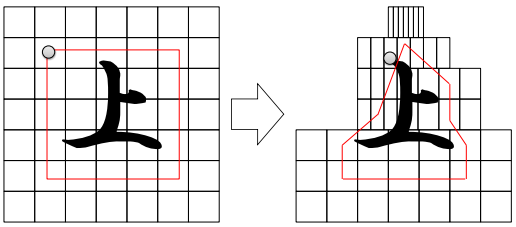
このように判断できる理由は，形状が三角に近い文字の場合，点を打つ位置が文字の形に合わせて変化するためである．「令」と「上」に関するヲコト点の位置を図 10 に示す．この図からもわかるように，図 9(a)の点が「令」の左上であるならば，もう少し文字に近接した位置になるはずである．

また，図 9(b)は，初めてヲコト点を読む者にとっては，解釈が難しい例である．物理的な位置は文字の中央上であるが，基本的にヲコト点は文字に近接させて打つという原則を考えると，この点が左上であることも納得できる．もちろんヲコト点の意味がわかる者にとっては，より位置の判断がしやすくなる．

最後の図 9(c)は，電子化時の処理が困難になりやすい例である．「下」の符合は文字の右のみに記入するように点図には記載されている．このため，機械的な変換に使う点図データには（2,0）の位置で登録されている．しかし実際の資料では文字の右側の適当な隙間に配置されていることが多く，システムの動作を知らなければ，入力者が（3,0）や（2,1）などの別の座標に登録してしまう可能性も高い．



(a) 「令」の場合



(b)「上」の場合

**図 10　文字の形によるグリッドの変化例**

### 5.2　対象とする利用者の漢文に対する理解度

今回，ヲコト点を電子的に記録し，表現するために入力支援システムが必要と考えた理由は，ヲコト点や点吐口訣など訓点が付与された資料は東アジア全体で広く存在しているものの，それらを読むことができる人が少ないため，誰でもこれらを簡単に読めるようにできないかと考えたためである．また，資料を電子化することで，ヲコト点の時代による変化や，他の訓点との関係をより詳しく分析できるようにしたいという目的もある．

しかしながら，現在までに製作したツールだけは，ヲコト点の知識がまったくない状態で使うことは難しい．特に，5.1 で述べた点の位置が文字にあわせて変化する問題に対処する必要がある．もちろん，誰が入力しても入力エラーは起こるものであるが，初心者が資料だけを参照して入力する際には，更なる補助が必要であると考えている．具体的には，書下し文がリアルタイムに表示され誤入力を判断しやすくする．文字によって頻出するヲコト点は決まっているので，専門家が入力したデータと比較し，統計的に候補を絞り込んで修正するなどの対応が必要である．

## 6．まとめと今後の予定

今回，ヲコト点の特徴を考慮した電子的に利用可能なデータ構造として，文字の中心を原点とする 7 × 7 の整数座標系を提案した．また，これに合わせた構造化記述のための XML フォーマットを検討した．

そして，このデータ構造を使用して，ヲコト点が付与された資料を扱うためのシステムを構築するために，ヲコト点情報付与ツールと，ヲコト点 Reader を製作した．

ヲコト点情報付与ツールを使ってヲコト点の入力を行ったところ，データ構造的には同じ座標を持つはずのヲコト点の物理的な位置が，文字の形状に合わせて移動してしまう点が，入力上の大きな課題であることが明らかとなった．

点図集による点図データの作成は完了したので，今後は，実際の資料の入力作業を継続すると共に，ヲコト点情報付与ツールの入力処理を初心者でも簡単にできるように改良を行う予定である．その後，書下し文を作成する以外の解析ツールの製作を行っていく予定である．

## 謝辞

本研究は国立国語研究所共同研究プロジェクト「訓点資料の構造化記述」による成果の一部です．

## 参考文献

[1] 築島裕：訓点語彙集成〈第 1 巻〉，ヲコト點概要, 汲古書院 (2007) .
[2] 朴 鎭浩：文字生活史の観点から見た口訣, 文学, 第 12 巻, 第 3 号, pp.169-181（2011）.