

国立国語研究所学術情報リポジトリ

ヲコト点の移点ツール

メタデータ	言語: jpn 出版者: 公開日: 2020-03-18 キーワード (Ja): キーワード (En): 作成者: メールアドレス: 所属:
URL	https://doi.org/10.15084/00002650

ヲコト点の移点ツール

田島 孝治

1 移転ツールの概要

漢文加点資料を電子化する場合、現在は原文のみ、または、ヲコト点を解釈した書下し文での記録が一般的である。この方法は、原文を読みやすく記録でき、資料の内容を重視する場合には適している。一方で、資料のどの位置にどんなヲコト点が打たれていたかの情報は記録されていないため、記録結果の解読において他の解釈を行うことが難しい。また、ヲコト点自体の分析を行う資料として利用できない問題もある。さらに、書下し文の作成には漢文訓読に関する専門的な知識が必要であり、書下し文の作成を行った人により記録結果が異なる可能性もある。

これらの問題を解決するには、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法を検討し、漢文加点資料に付与されたヲコト点に関する情報を電子的に記録するツールが必要である。ヲコト点に関する情報とは、原文に付与された点の位置、形状、色などを表す。本稿では、このツールのことを「移点ツール」と呼ぶことにする。「移点」とは、元来ヲコト点を事前に用意した紙媒体などに写し取る作業を表すが、ここでは電子的に写し取るという意味で同じ言葉を用いている。

移点ツールを利用したヲコト点資料の電子化の流れを図1に示す。初めに資料から文字だけを抜き出しテキストデータを作成する。このデータを漢文テキストと呼ぶ。次に、移点ツールを使い、資料を見ながら漢文テキストにヲコト点情報を追加する。移点ツールは、この結果を構造化テキストとして出力する。移点ツールを使うことで、(1)記述が大変なヲコト点情報を直感的に記録できる、(2)ヲコト点情報を付与した構造化テキストの体裁が一意になり利活用が容易になる、という効果が期待できる。

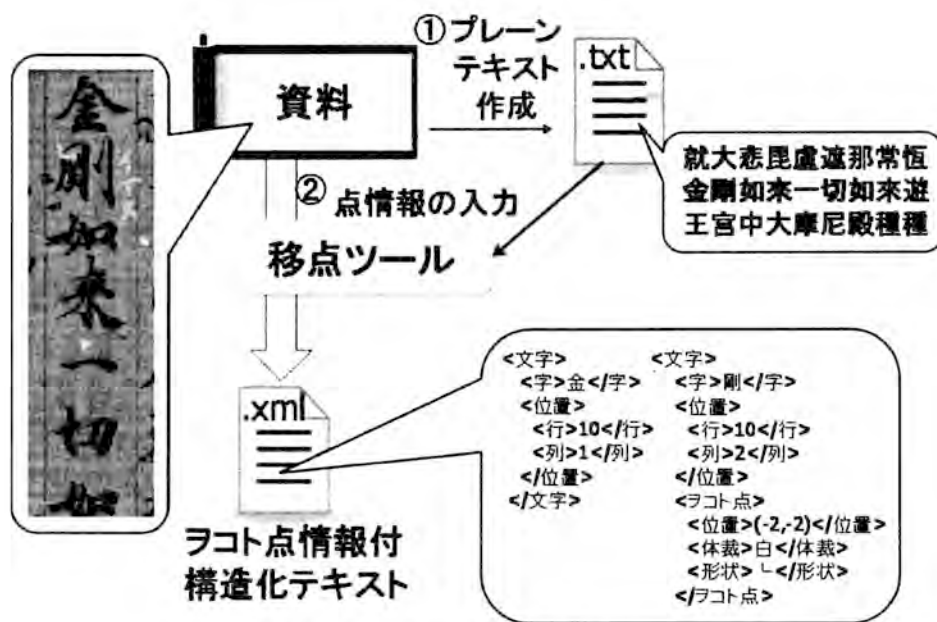


図1 ヲコト点情報を付与した構造化データの作成

2 入出力データの仕様

(1) 漢文テキスト

漢文テキストの形式はプレーンテキストとする。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ため、手作業での入力を想定した書式が求められる。このため、タグを入れ子にするような複雑な記述方法は適していない。また、これまでの研究の中で、訓点資料の白文のテキストデータが既に制作されているものもあることを考えると、特別な構造を表す記述を追加することは好ましくない。

一方で、資料のテキストデータが何の区切りもなく記述されていると、ヲコト点を移点する作業において位置の判断が困難になる。そこで、漢文テキストにおいては、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）と、テキストデータ中の論理的な行数（以後、論理行と呼ぶ）を一致させておく必要がある。

別の問題として、訓点資料には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、プレーンテキストとして入力できない文字（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。移点ツールにはこれらの簡易タグを解釈し、物理行のデザインに合わせた表示を行う機能が求められることになる。

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]

(2) フォント情報付与済み構造化テキスト

フォント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。フォン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。

XMLの構造を図2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)フォント点を付与する。フォン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」→「フォント点」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。フォント点は、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。

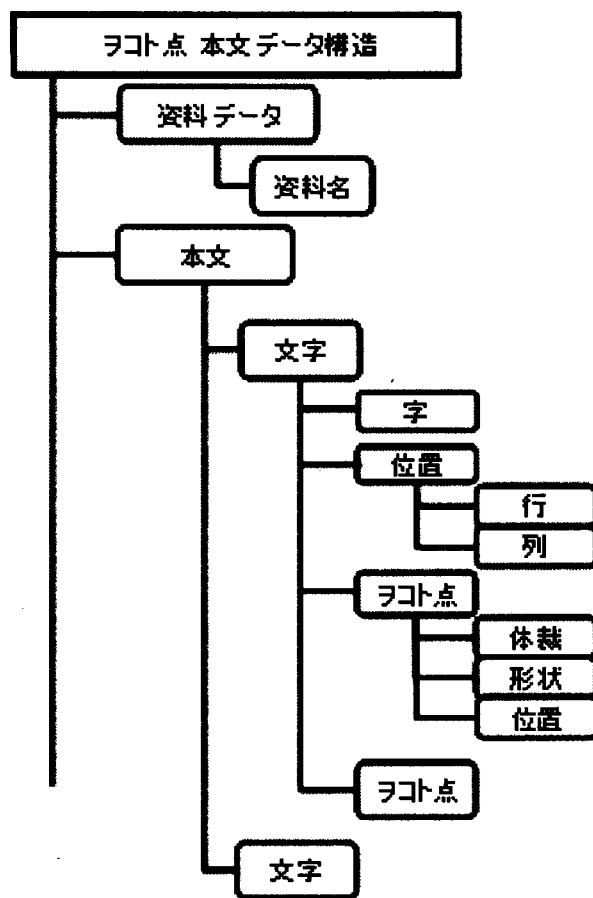


図2 ヲコト点情報付与済みテキストの構造

3 移点ツール

移点ツールの GUI を図3、図4に示す。図3は基本画面で、ツールを使いプレーンテキストを開いた状態である。図4は一文字に注目しヲコト点を転記中の画面である。

基本画面における「Open」のメニューには、「テキストファイルを開く」、「文字コードを選択して開く」、「タグを無視して開く」、「XMLを開く」の4種類が格納されている。「テキストファイルを開く」を選択すると、簡易タグを処理しながら UTF-8 でエンコードされたテキストファイルを読み込むことができる。他の文字エンコードを利用している場合には「文字コードを選択して開く」の機能を使う。この機能は、基本画面下部のボタンからも呼び出すことができ、操作性を高めている。「タグを無視して開く」は簡易タグをタグとして認識せずに開く機能であり、タグにより表示に不都合が生じた際などに利用する。「XMLを開く」は、保存した作業結果を開き、編集や追加を行うための機能である。こちらも、画面下部のボタンで同様の機能を呼び出すことができる。

基本画面における「Save」メニューは、「名前を付けて保存」と「上書き保存」の機能を呼び出すことができ、それぞれ移点結果を XML として保存するために利用する。

基本画面の中央には、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ことで、点情報を入力するためのダイアログが開く。



図3 移転ツールの基本画面（テキストを開いた直後の様子）

基本画面の文字をクリックすることで表示された点情報を入力するためのダイアログが図4である。ツールの利用者は、資料を見ながら、この画面を利用してヲコト点の移点作業を行う。

移点作業は、ヲコト点一点の移点は次の3つの手順で行う。(1)何の筆記具を用いたかを表す「点の体裁」を選択する。(2)どんな形状の点が打ってあるかを「点の形状」から選択する。(3)点の打たれている場所を、左側の文字上のグリッドから選択する。ある文字に記述されたヲコト点の数だけこの手順を繰り返すことで、その文字の移点作業は完了する。一つの文字にどんな点を打ったかは、画面の右側に一覧表で表示され、ここから削除もできるようになっている。

一文字の移点作業が完了したら、「次の文字」ボタンを押すことで次の文字に対する移点作業が行える。なお、資料に同じ文字列が繰り返し登場した際に、場所を間違えにくくするために、作業中の文字がどこにあるのかが基本画面に表示されるよう工夫した。

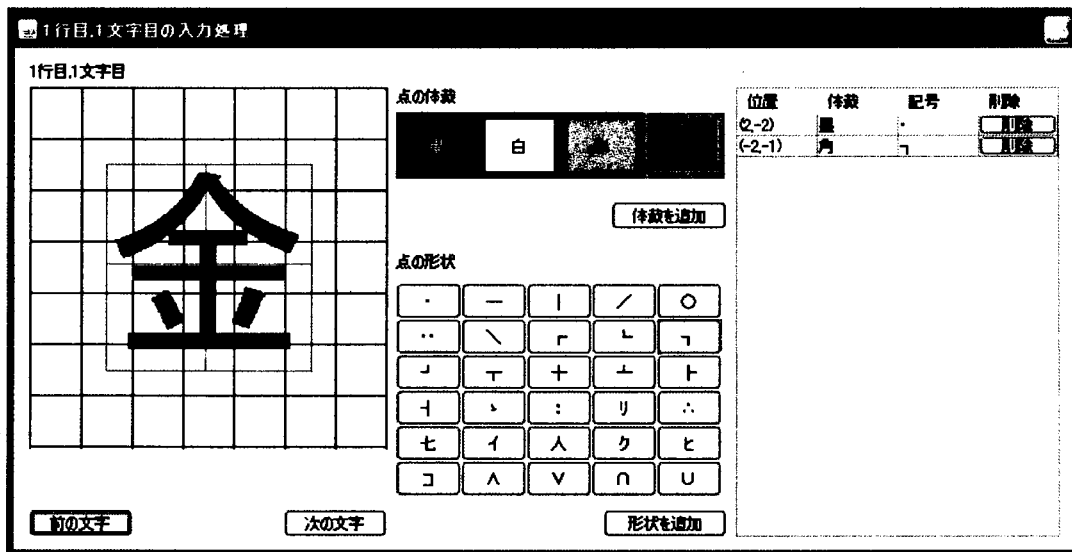


図4 フコト点転記用の画面

4 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。この写本は喜多院点のフコト点が主に白点で付与されている。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ので、ここにまとめる。

4. 1 文字の形状による点の位置変化

移点時に一番問題となったのは、文字の形に合わせて点の位置が変化することである。これは、「剛」のような輪郭が四角形に近い文字であれば起こらない問題である。問題が起こるのは「一」、「入」、「上」、「中」のような、字画の形状が極端に横長の文字または、三角形の文字である。

図5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フコト点の例を示す。(a)は、点がどちらに付与されているのかわ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付きにくい。一方で、(b)の点は「无(無)」下辺中央と見える可能性は少ないが、「上」の上辺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、その関係がわかりにくい。

これらの解釈は、フコト点が加えら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。現在の解釈では、(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、点を打つ位置が文字の形に合わせて変化する原理による。

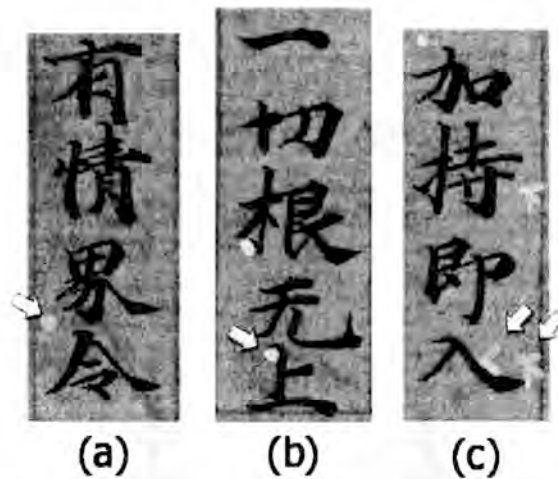


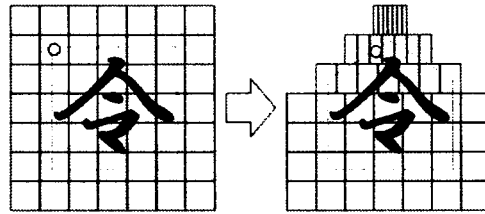
図5 文字の形によるマコト点の位置のずれ
(金剛頂一切如來眞實攝大乘現證大教王經より)

「令」の左上に点を打つ場合には、図6 (a)のように、令の上辺中央あたりに点が打たれることが一般的である。「上」に関しても同様に左上の点は、図6 (b)のような位置になる。どちらもグリッドは綺麗な正方形にならず、文字の形に合わせて変形してしまっている。マコト点の解読に慣れ、点の位置から意味を解釈できる作業者であれば、この点の位置を正しく入力できるが、何も分からない状態では難しい。

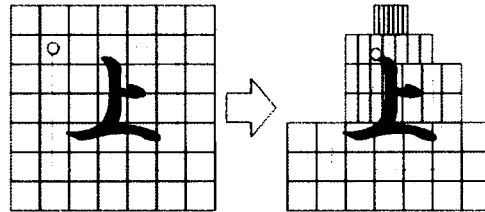
図5 (c)は、経験がなければ、正しく電子化できない可能性が高い例である。「下」の符合は文字の右側に記入するように点図に記載されている。このため、入力される座標として(2, 0)が適切である。しかし実際の資料では文字の右側の適当な隙間に配置されていることが多く、入力者が(3, 0)や(2, 1)などの別の座標に登録してしまう可能性が高い。

4. 2 特殊な意味や位置を持つ訓点の処理方法

もう一つ問題となったのは、一つの文字だけに関係するわけではない記号の処理である。具体的には、句読点と合符について考える必要がある。これら二つは本質的にはマコト点ではないため、本ツールで入力する必要はないが、その形状がマコト点と同一、または類似であるため、入力者にこれらの違いを判断させる必要がでてくる。また、構造化テキストにおいて句読点や合符が抜けていると、例えば書下し文の作成に利用しようと考えた際に、単語を正しく読み取れない可能性もある。



(a)「令」の場合



(b)「上」の場合

図6 文字の形に合わせたグリッド形状の変更

5 まとめ

本稿では、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法を検討し、これらの資料に付与されたヲコト点に関する情報を電子的に記録する移点ツールの開発について述べた。本ツールを利用した漢文加点資料の移点作業は、専用の GUI を用いるため、ヲコト点の色や形状を選び、画面上に表示された文字上の加点位置をクリックするだけで実現できる。また、この結果は XML の汎用的な形式で保存され再利用が容易である。

今後は、漢文加点資料に関する知識がないユーザの利用を考慮した、文字の形状に合わせ点の位置が変化する問題への対応や、句読点を表す点や合符などを扱うための改良の実施、さらには、移点作業を行ったことがある研究者に試用してもらい、ツールの不足機能を明らかにしていく予定である。

参考文献

- 吉田金彦・築島裕・石塚晴通・月本雅幸 (2001) 『訓点語辞典』東京堂出版
 朴鎮浩 (2011) 「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
 田島孝治・堤智昭・高田智和 (2012) 「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論文集 つながるデジタルアーカイブ一分野・組織・地域を越えて』、pp.211-216